

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2005, 08:43:45 ; Search time 40 Seconds  
(without alignments)  
1062.746 Million cell updates/sec

Title: US-10-757-262-104  
Perfect score: 3480  
Sequence: 1 MPPPPPLLLLTVLVVAARP.....WWSYFTSLSTHRPRLKY 641

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3480	100.0	641	3	US-09-233-989-10
2	1057	30.4	476	3	US-09-233-989-3
3	1054	30.3	476	3	US-09-233-989-2
4	1054	30.3	476	4	US-09-917-254-67
5	1052	30.2	434	1	US-08-111-939-13
6	1052	30.2	434	3	US-09-233-989-7
7	1051	30.2	435	1	US-08-111-939-16
8	1051	30.2	435	1	US-08-452-262-2
9	1051	30.2	435	1	US-08-734-550-2
10	1051	30.2	435	5	PCT-US96-07528-2
11	1050	30.2	435	1	US-08-111-939-14
12	1049	30.1	476	3	US-09-233-989-6
13	1047	30.1	435	1	US-08-111-939-15
14	1036	29.8	734	4	US-09-641-741-2
15	1035	29.7	454	3	US-09-233-989-4
16	1032	29.7	734	3	US-08-706-216-2
17	1032	29.7	734	4	US-09-650-284B-2
18	1003	28.8	722	4	US-09-641-741-32
19	1001	28.8	438	1	US-08-111-939-17
20	1001	28.8	438	3	US-09-233-989-9
21	961	27.6	756	4	US-10-140-002-392
22	944	27.1	506	4	US-09-370-838-34
23	944	27.1	506	4	US-09-854-133-34
24	944	27.1	764	4	US-09-641-741-31
25	906.5	26.0	1128	1	US-08-111-939-2
26	906.5	26.0	1128	4	US-09-641-741-30
27	904.5	26.0	719	4	US-09-641-741-28

28	903.5	26.0	484	1	US-08-111-939-12	Sequence 12, Appl
29	898.5	25.8	1128	4	US-09-060-482-8	Sequence 8, Appli
30	895.5	25.7	845	4	US-09-641-741-29	Sequence 29, Appl
31	895.5	25.7	1158	4	US-09-060-482-2	Sequence 2, Appli
32	867.5	24.9	377	4	US-09-148-545-140	Sequence 140, App
33	771	22.2	561	3	US-09-233-989-5	Sequence 5, Appli
34	764.5	22.0	372	2	US-08-683-262B-64	Sequence 64, Appl
35	764.5	22.0	372	3	US-09-361-707-64	Sequence 64, Appl
36	707	20.3	439	3	US-09-233-989-8	Sequence 8, Appli
37	707	20.3	443	4	US-09-976-594-527	Sequence 527, App
38	511.5	14.7	208	4	US-09-148-545-207	Sequence 207, App
39	209.5	6.0	281	3	US-08-893-654B-2	Sequence 2, Appli
40	200.5	5.8	280	3	US-08-893-654B-4	Sequence 4, Appli
41	195	5.6	295	3	US-08-893-654B-6	Sequence 6, Appli
42	183.5	5.3	565	4	US-08-937-067-8	Sequence 8, Appli
43	180.5	5.2	585	4	US-08-937-067-9	Sequence 9, Appli
44	175	5.0	317	4	US-08-937-067-6	Sequence 6, Appli
45	173.5	5.0	666	4	US-08-937-067-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-09-233-989-10  
; Sequence 10, Application US/09233989  
; Patent No. 6248527  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Meyer, Joanne  
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on  
; FILE OF INVENTION: Mutations Found in Carboxypeptidase E  
; FILE REFERENCE: 5800-14, 035800/174130  
; CURRENT APPLICATION NUMBER: US/09/233,989  
; CURRENT FILING DATE: 1999-01-19  
; EARLIER APPLICATION NUMBER: 60/105,102  
; EARLIER FILING DATE: 1998-10-21  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 641  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: carboxypeptidase homolog -- CPZ  
US-09-233-989-10

Query Match	100.0%;	Score	3480;	DB	3;	Length	641;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	641;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MPPPPPLLLLTVLVVAARP	CGCEFFERNPAATCVDLQ	LTCTSDAAYNHHTFPNLLQ	HRSW	60	
Db	1	MPPPPPLLLLTVLVVAARP	CGCEFFERNPAATCVDLQ	LTCTSDAAYNHHTFPNLLQ	HRSW	60	
QY	61	VVEASSEYILLSVLHOLLEG	QCNPDLRLGCAVLAPRC	EGGWVRRPCRIHCEGL	REVCQ	120	
Db	61	VVEASSEYILLSVLHOLLEG	QCNPDLRLGCAVLAPRC	EGGWVRRPCRIHCEGL	REVCQ	120	
QY	121	AFDAIDMAWPFILDC	HRYTREDEGCYDPLEK	RGLEADEALPSGLPTT	FIHFSSH	180	
Db	121	AFDAIDMAWPFILDC	HRYTREDEGCYDPLEK	RGLEADEALPSGLPTT	FIHFSSH	180	
QY	181	QMVRLRRRTASCAH	VARTYSICRSPDGR	ELLVIESSRPGQHELM	EPVKLLIGN	240	
Db	181	QMVRLRRRTASCAH	VARTYSICRSPDGR	ELLVIESSRPGQHELM	EPVKLLIGN	240	
QY	241	VAGREMLIYLAQYLC	SEYLLGNPRIORLNT	TRIHLPSINPDGYE	VAAAGAGY	300	
Db	241	VAGREMLIYLAQYLC	SEYLLGNPRIORLNT	TRIHLPSINPDGYE	VAAAGAGY	300	
QY	301	GRQNAQLNLRN	FPDLTSEYVRLAET	RGARDSHIPQHY	WMGKVP	360	
Db	301	GRQNAQLNLRN	FPDLTSEYVRLAET	RGARDSHIPQHY	WMGKVP	360	

Db 301 GRQNAQLDLNRNPFDLTSEYRLAETRGARSDHIPIPOHYWGWKVAPETKAIKMWQTI 360  
 QY 361 PFVLSASLHGDLVVSYPDFSKHPOEKFSPTPDEKMFKLLSRAYADVHPMMDRSEN 420  
 Db 361 PFVLSASLHGDLVVSYPDFSKHPOEKFSPTPDEKMFKLLSRAYADVHPMMDRSEN 420  
 QY 421 RCGNFKLRGSIINGADWYSGTGMSPFNLYHTNCPETITVELGCVKFPPEEALYTLWOHN 480  
 Db 421 RCGNFKLRGSIINGADWYSGTGMSPFNLYHTNCPETITVELGCVKFPPEEALYTLWOHN 480  
 QY 481 KESLLNFVETVHRGIGKGVWTDKFGKPKVNARIISVKGIRHDIITAPDGDYWRLLPPGHIHV 540  
 Db 481 KESLLNFVETVHRGIGKGVWTDKFGKPKVNARIISVKGIRHDIITAPDGDYWRLLPPGHIHV 540  
 QY 541 IAQAPGVAKVKKVILPARKRAGRVDPILOPLMGPKNFHGLRRTGPHDPIGGASSLG 600  
 Db 541 IAQAPGVAKVKKVILPARKRAGRVDPILOPLMGPKNFHGLRRTGPHDPIGGASSLG 600  
 QY 601 EATEPDLRARROPSADGSPWWSYFTSLSTRPRLWKY 641  
 Db 601 EATEPDLRARROPSADGSPWWSYFTSLSTRPRLWKY 641

## RESULT 2

US-09-233-989-3

; Sequence 3, Application US/092333989

; Patent No. 6248527

; GENERAL INFORMATION:

; APPLICANT: Chen, Hong

; APPLICANT: Meyer, Joanne

; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on

; TITLE OF INVENTION: Mutations Found in Carboxypeptidase E

; FILE REFERENCE: 5800-14, 035800/174130

; CURRENT APPLICATION NUMBER: US/09/233,989

; CURRENT FILING DATE: 1999-01-19

; EARLIER APPLICATION NUMBER: 60/105,102

; EARLIER FILING DATE: 1998-10-21

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 3

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Rattus sp.

; FEATURE:

; OTHER INFORMATION: carboxypeptidase E

US-09-233-989-3

Query Match 30.4%; Score 1057; DB 3; Length 476;  
 Best Local Similarity 48.4%; Pred. No. 6.7e-97;  
 Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHARTYSI 202  
 Db 20 GMLLAQAEPGAPAAAGMRRLRRLQOEDGISFEYHRYPELREALVSVMLQCTAISRIYTV 79  
 QY 203 GRSGDRELLVIEFSSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEVLLGN 262  
 Db 80 GRSGFEGRELLVIELSDNPGVHEGPEFVKYIGNHNEAVGRELLIFLAQYLCNEYQKGN 139  
 QY 263 PRIQRLNLTTRIHLPSINPDGYEVAAGAGYNGWTSGRQNAQLDLNRNPPDLTSEY 322  
 Db 140 ETIVNLHSTRIHIMPSLNPDPGEKAAQSGELKDFVGRSNAQGLDLNRNPPDLDRIVY 199  
 QY 323 RLAEETRGARSDHI-----PIPOHYWGWKVAPETKAIKMWQTIIPVLSASLHGDLVVS 376  
 Db 200 -VNEKEGGNNHLLKMKKIVDQN---SKLAPETKAVIHWIMDIPFVLSANLHGDLVAN 255  
 QY 377 YPFDPSKHPOEKFSPTPDEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
 Db 256 YPYDETRSGTAHE-YSSCPDDAIFQSLARAYSFNPMVSDPNRPPCKNDDSSFV--DG 312  
 QY 432 IINGADWYSGTGMSPFNLYHTNCPETITVELGCVKFPPEEALYTLWOHNKESLLNFVETV 491  
 Db 432 IINGADWYSGTGMSPFNLYHTNCPETITVELGCVKFPPEEALYTLWOHNKESLLNFVETV 491

Db 313 TTNGGAWYSPVGMQDFNYLSSNCPETITVELSCEKPPPEETLKSYWEDKNKSLINYLEQI 372  
 QY 492 HRGIGKGVWTDKFGKPKVNARIISVKGIRHDIITAPDGDYWRLLPPGHIHIVIAQAPGVAKY 551  
 Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKADGDYWRLLVPGNYKLTASAPGYLAIT 432  
 QY 552 KKVILPARKRAGRVDPILO 571  
 Db 433 KKVAVP--FSPAAGVDVFELE 450

## RESULT 3

US-09-233-989-2

; Sequence 2, Application US/092333989

; Patent No. 6248527

; GENERAL INFORMATION:

; APPLICANT: Chen, Hong

; APPLICANT: Meyer, Joanne

; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on

; TITLE OF INVENTION: Mutations Found in Carboxypeptidase E

; FILE REFERENCE: 5800-14, 035800/174130

; CURRENT APPLICATION NUMBER: US/09/233,989

; CURRENT FILING DATE: 1999-01-19

; EARLIER APPLICATION NUMBER: 60/105,102

; EARLIER FILING DATE: 1998-10-21

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: Patent in Ver. 2.0

; SEQ ID NO 2

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-233-989-2

Query Match 30.3%; Score 1054; DB 3; Length 476;  
 Best Local Similarity 48.2%; Pred. No. 1.3e-96;  
 Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHARTYSI 202  
 Db 20 GMLLAQAEPGAPAAAGMRRLRRLQOEDGISFEYHRYPELREALVSVMLQCTAISRIYTV 79  
 QY 203 GRSGDRELLVIEFSSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEVLLGN 262  
 Db 80 GRSGFEGRELLVIELSDNPGVHEGPEFVKYIGNHNEAVGRELLIFLAQYLCNEYQKGN 139  
 QY 263 PRIQRLNLTTRIHLPSINPDGYEVAAGAGYNGWTSGRQNAQLDLNRNPPDLTSEY 322  
 Db 140 ETIVNLHSTRIHIMPSLNPDPGEKAAQSGELKDFVGRSNAQGLDLNRNPPDLDRIVY 199  
 QY 323 RLAEETRGARSDHI-----PIPOHYWGWKVAPETKAIKMWQTIIPVLSASLHGDLVVS 376  
 Db 200 -VNEKEGGNNHLLKMKKIVDQN---TKLAPETKAVIHWIMDIPFVLSANLHGDLVAN 255  
 QY 377 YPFDPSKHPOEKFSPTPDEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
 Db 256 YPYDETRSGTAHE-YSSCPDDAIFQSLARAYSFNPMVSDPNRPPCKNDDSSFV--DG 312  
 QY 432 IINGADWYSGTGMSPFNLYHTNCPETITVELGCVKFPPEEALYTLWOHNKESLLNFVETV 491  
 Db 313 TTNGGAWYSPVGMQDFNYLSSNCPETITVELSCEKPPPEETLKSYWEDKNKSLINYLEQI 372  
 QY 492 HRGIGKGVWTDKFGKPKVNARIISVKGIRHDIITAPDGDYWRLLPPGHIHIVIAQAPGVAKY 551  
 Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKADGDYWRLLVPGNYKLTASAPGYLAIT 432  
 QY 552 KKVILPARKRAGRVDPILO 571  
 Db 433 KKVAVP--YSPAAGVDVFELE 450

## RESULT 4

US-09-917-254-67

; Sequence 67, Application US/09917254

```
; Patent No. 6703204
; GENERAL INFORMATION:
; APPLICANT: Mutter, George
; APPLICANT: Baak, Jan
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer
; FILE REFERENCE: B0801/7224(JRV)
; CURRENT APPLICATION NUMBER: US/09/917,254
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: US 60/222,093
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-09-917-254-67

Query Match      30.3%; Score 1054; DB 4; Length 476;
Best Local Similarity 48.2%; Pred. No. 1.3e-96;
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GLEADEALPSGLPPTFIR-----FSHHSYAQMVRVLRTASTCAHVARTYSI 202
DB 20 GWLGAQAQPGAPAGMRRRLQQDDGISFHYRYPELRALVSVMLQCTAISRIYTV 79

QY 203 GRSPDGRLLVIEFSSRPGQHELMPEVKLIGNHNEVAGREMLIYLAQYLCSEYLLGN 262
DB 80 GRSPGRELIVIELSDNPGVHEPEPEFKYIGNMHGNEAVGRELLIIFLAQYLCNEYQGN 139

QY 263 PRQRLNTHRIHLPLINPDGVEVAAAGAGYNGWTSGRQNAQLDLNRPFDLTSEYY 322
DB 140 ETIVNLHSTRIHIMPISLNPDEGFEKAASQPCGLKDFWFGRSNAQGLDLNRPFDLRIYV 199

QY 323 RLAEATRCARSDHI-----PIPOHYWGWKVPETKATMKMOTIPFVLSASLHGGDLVVS 376
DB 200 -VNEKEGGPNHLLKNMKIVDQ---TKLAPETKAVIHWIMDIPFVLSANLHGGDLVAN 255

QY 377 YPDPFSKHPQBEKMFSTPDEKMPKLSRAYADVHPMMDRSENRCGN-----FLKRG 431
DB 256 YPDETRSGSAHE-YSSSPDAIFQSLARAYSFNPAMSPNPPCRKNDSDSFV--DG 312

QY 432 IINGADWYFTGGMDSFNYLHTRCFEITVELGCVKFPPEALYTLWOHNKESLINFVETV 491
DB 313 TTGGAWYSPVGGMDFNLSNCFEITVELSCEKFPPEETLKYWNEDNKNLSIYLEQI 372

QY 492 HRGKGVVTKFGKPVKNARISVKGIRHDITAPDGYWRLPPGIHIVIAQAPYAKVI 551
DB 373 HRGKGVFVRBLOQNPANATISVEGIDHDVTSKQDGYWRLPIPGNYKLTAAPGYLAIT 432

QY 552 KKVIIIPARMKRAGRVDFILQ 571
DB 433 KKVAVP--YSPAAGVDFELE 450

RESULT 5
US-08-111-939-13
; Sequence 13, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
; APPLICANT: Kawai, Shinji
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
; TITLE OF INVENTION: Protein and Process for its Production
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
```

```
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/111,939
; FILING DATE: 26-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324033/92
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 230029/92
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1321-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4000
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-111-939-13

Query Match      30.2%; Score 1052; DB 1; Length 434;
Best Local Similarity 50.1%; Pred. No. 1.8e-96;
Matches 206; Conservative 67; Mismatches 118; Indels 20; Gaps 7;

QY 172 IRFSHHSYAQMVRVLRTASTCAHVARTYSIGSPDGRLLVIEFSSRPGQHELMPEVK 231
DB 7 ISPEYHRYPELRALVSVMLQCAAVSRITYVGRSFEGLLELLELSNPGVHEPGSPFEK 66

QY 232 LIGNHNEVAGREMLIYLAQYLCSEYLLGNPRIQLNTHRIHLPLINPDGVEVAAAE 291
DB 67 YIGNMHGNEAVGRELLIIFLAQYLCNEYQGNETIVQLIHNTHRIHIMPISLNPDEGFEKAASQ 126

QY 292 GAGYNGWTSGRQNAQLDLNRPFDLTSEYVRLAETRGARSDHI-----PIPOHYWGWK 345
DB 127 LGELKDFWFGRSNAQGLDLNRPFDLRIYV-INEKEGGNNHLLKNMKIVDQ---TK 182

QY 346 VAPETKAIKMKMOTIPFVLSASLHGGDLVVSYPDFSKHPQBEKMFSTPDEKMPKLSR 405
DB 183 LAPETKAVIHWIMDIPFVLSANLHGGDLVANYFYDETRSGSAHE-YSSCPDDDDIFQSLAR 241

QY 406 AYADVHPMMDRSENRCGN-----FLKRSIINGADWYFTGGMDSFNYLHTRCFEITV 460
DB 242 AYSSFNPPSPDPRPPCRKNDSDSFVE--GTTNGAAWYSPVGGMDFNLSNCFEITV 299

QY 461 ELGCVKFPPEALYTLWOHNKESLINFVETVHRGIRGVVTKFGKPVKNARISVKGIRHD 520
DB 300 ELSCERFPPEETLKNYWNEDNKNLSISYIQIHGVGVGFVRDLQGNPANNATLSVEGIDHD 359

QY 521 ITTAPDGYWRLPPGIHIVIAQAPYAKVIKVIIPARMKRAGRVDFILQ 571
DB 360 VTSKQDGYWRLPIPGNYKLTAAPGYLAITAKKAVP--YSPAVRVDFELE 408

RESULT 6
US-09-233-989-7
; Sequence 7, Application US/09233989
; Patent No. 6248527
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Meyer, Joanne
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on
```

```
; TITLE OF INVENTION: Mutations Found in Carboxypeptidase E
; FILE REFERENCE: 5800-14, 035800/174130
; CURRENT APPLICATION NUMBER: US/09/233,989
; CURRENT FILING DATE: 1999-01-19
; EARLIER APPLICATION NUMBER: 60/105,102
; EARLIER FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 434
; TYPE: PRT
; ORGANISM: alysia
; FEATURE:
; OTHER INFORMATION: carboxypeptidase E
US-09-233-989-7

Query Match          30.2%; Score 1052; DB 3; Length 434;
Best Local Similarity 50.1%; Pred. No. 1.8e-96;
Matches 206; Conservative 67; Mismatches 118; Indels 20; Gaps 7;

QY 172 IRFSHSHYAQMVRVLRRTASRCARHARTVYSIGRSPDRELLVIEFSSRPGQHELMEPEVK 231
DB 7 ISPEYHRYPELREALVSVWLQCAAVSRITYTVGRSFEGLLELVLELSDNPGVHEPGEPEFK 66

QY 232 LIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIQRLNTRIHLLPSINPDGYEVAAB 291
DB 67 YICNMHGNEAVGRELIFLAQYLCNEYOKNETIVQLIHNTRIHIMPSLNPDGFEKAAQ 126

QY 292 GAGYNGWTSGRQAQNLNLRNFPDLTSEYRRLAETRGARSDHI-----PIPOHYWVGK 345
DB 127 LGBELKDFWFGSRNAQIGIDLNRNFPDLRIYV-INEKEGGPNHLLKNKKIVDQN---TK 182

QY 346 VAPETKAIKMKWQTIPIFVLSASLHGGDLVSVYFDFSKHPQEKMFSPPTDEKMFKLISR 405
DB 183 LAPETKAVIHWMIDIPFVLSANLHGGDLVANYPYDETRGSAHE-YSSCPDDDDIFOSLAR 241

QY 406 AYADVHPMMMDRSENRCGN-----FLKRGSIINGADWYSFTGCMGDFNYLHTNCFEITV 460
DB 242 AYSSFPNMSDPNRPCKRKNDDSSFVE--GTITNGAAYSVPGQMDFNLSNCFEITV 299

QY 461 ELGCVKFPPEALYTLWQHNKESLLNFVETVHRGKGVVTDKFGPKVKARISVKGIRHD 520
DB 300 ELSCFKFPPEETLKYTWEDNKNLSIYQIHRGKGVFVRDLQGNPIANATLSVEGIDHD 359

QY 521 ITTAPDGYWRLLPPGIHIVIAQAPGYAKVIKKVILPARKMKRAGRVDFILQ 571
DB 360 VTSKDGWYRLLVPNGYKLTASAPGYLAITAKKVAVP--YSPAVRVDFELE 408

RESULT 7
US-08-111-939-16
; Sequence 16, Application US/08111939
; Patent No. 5460951
; GENERAL INFORMATION:
; APPLICANT: Kawai, Shinji
; APPLICANT: Takeshita, Sunao
; APPLICANT: Okazaki, Makoto
; APPLICANT: Anann, Egon
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like
; TITLE OF INVENTION: Protein and Process for its Production
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/111,939
; FILING DATE: 26-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324033/92
; FILING DATE: 03-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 230029/92
; FILING DATE: 28-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1321-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4000
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-111-939-16

Query Match          30.2%; Score 1051; DB 1; Length 435;
Best Local Similarity 50.1%; Pred. No. 2.3e-96;
Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;

QY 172 IRFSHSHYAQMVRVLRRTASRCARHARTVYSIGRSPDRELLVIEFSSRPGQHELMEPEVK 231
DB 8 ISFEYHRYPELREALVSVWLQCAVSRITYTVGRSFEGLLELVLELSDNPGVHEPGEPEFK 67

QY 232 LIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIQRLNTRIHLLPSINPDGYEVAAB 291
DB 68 YICNMHGNEAVGRELIFLAQYLCNEYOKNETIVQLIHNTRIHIMPSLNPDGFEKAAQ 127

QY 292 GAGYNGWTSGRQAQNLNLRNFPDLTSEYRRLAETRGARSDHI-----PIPOHYWVGK 345
DB 128 PGELKDFWFGSRNAQIGIDLNRNFPDLRIYV-VNEKEGGPNHLLKNKKIVDQN---TK 183

QY 346 VAPETKAIKMKWQTIPIFVLSASLHGGDLVSVYFDFSKHPQEKMFSPPTDEKMFKLISR 405
DB 184 LAPETKAVIHWMIDIPFVLSANLHGGDLVANYPYDETRGSAHE-YSSFPDDDAIFOSLAR 242

QY 406 AYADVHPMMMDRSENRCGN-----FLKRGSIINGADWYSFTGCMGDFNYLHTNCFEITV 460
DB 243 AYSSFPNMSDPNRPCKRKNDDSSFV--DGTITNGAAYSVPGQMDFNLSNCFEITV 300

QY 461 ELGCVKFPPEALYTLWQHNKESLLNFVETVHRGKGVVTDKFGPKVKARISVKGIRHD 520
DB 301 ELSCFKFPPEETLKYTWEDNKNLSIYQIHRGKGVFVRDLQGNPIANATLSVEGIDHD 360

QY 521 ITTAPDGYWRLLPPGIHIVIAQAPGYAKVIKKVILPARKMKRAGRVDFILQ 571
DB 361 VTSKDGWYRLLVPNGYKLTASAPGYLAITAKKVAVP--YSPAAGVDFELE 409

RESULT 8
US-08-452-262-2
; Sequence 2, Application US/08452262
; Patent No. 553837
; GENERAL INFORMATION:
; APPLICANT: Naggett, Jurgen K.
; APPLICANT: Leitter, Edward H.
; TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH
; TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kevin M. Farrell, P.C.
; STREET: P.O. Box 999
; CITY: York Harbor
```

STATE: ME  
COUNTRY: US  
ZIP: 03911  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/452,262  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Farrell, Kevin M.  
REGISTRATION NUMBER: 35,505  
REFERENCE/DOCKET NUMBER: JL-9501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (207) 363-0558  
TELEFAX: (207) 363-0528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-452-262-2

Query Match 30.2%; Score 1051; DB 1; Length 435;  
Best Local Similarity 50.1%; Pred. No. 2.3e-96;  
Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;  
QY 172 IREHSHSYAQMVRVLRRTASRCARHARTYSGRFDGRELIVIEFSSRPGQHLMPEVK 231  
DB 8 ISFEYHRYPELRREALVSVMLQCTAISRIYTVGRSFEGRLLVIELSDNPVGHEPEFK 67  
QY 232 LIGNIHGNEVAGREMLIYLAQYLCEYLLGNPRIQRLNTRIHLLPSINPDGYEVA 291  
DB 68 YIGMHGNEAVGRELLIFLAQYLCNEYQKGNETIVNLHSTRIHIMPSLNPDGFEKASQ 127  
QY 292 GAGYNGWTSQRQNAQNLNRPDLTSEYVRLAETGARS DHI-----PIQHYWVGK 345  
DB 128 PGLKQWVGRSNAQGLDLNRPDLDRIVY-VNEKEGGNNHLLKNMKIVDQN---TK 183  
QY 346 VAPETKAIMKWMQTIPIFVLSASLHGGDLVSYPPDFSKHPQEEKMFSPPTDEKMFLLSR 405  
DB 184 LAPETKAVIHWINDIPFVLSANLHGGDLVANYPYDETSGSAHE-YSSSPDDAIFQSLAR 242  
QY 406 AYADVHPMMDRSENRCGN-----FLKRGSIINGADWYSFTGMSDFNYLHNCFEITV 460  
DB 243 AYSSFNPMSPDNPFPCKRKNDDSSFV--DGTNGGAWYSVPGMQDFNYLSSNCFEITV 300  
QY 461 ELGCVKFPPEEALYTLQHNKESLLNFVETVHRGKGVVTDKFGKPKVNARI SVKGRHD 520  
DB 301 ELSCEKFPPEETLKTWEDNKNLSIYLEQIHGRGVKGFVRDLQGNPIANATISVEGIDHD 360  
QY 521 ITTAPDGDYWRLLPPGHIHIVIAQAPGKAVIKKVIIPARKMRAGRVDFILO 571  
DB 361 VTSKAGDGYWRLLIPGNYKLTASAPGYLATKKVAVP--YSPAAGVDFELE 409

RESULT 9  
US-08-734-550-2  
Sequence 2, Application US/08734550  
Patent No. 5690932  
GENERAL INFORMATION:  
APPLICANT: Naggett, Jurgen K.  
APPLICANT: Leiter, Edward H.  
TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH  
TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kevin M. Farrell, P.C.  
STREET: P.O. Box 999

CITY: York Harbor  
STATE: ME  
COUNTRY: US  
ZIP: 03911  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/734,550  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Farrell, Kevin M.  
REGISTRATION NUMBER: 35,505  
REFERENCE/DOCKET NUMBER: JL-9501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (207) 363-0558  
TELEFAX: (207) 363-0528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-734-550-2  
Query Match 30.2%; Score 1051; DB 1; Length 435;  
Best Local Similarity 50.1%; Pred. No. 2.3e-96;  
Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;  
QY 172 IREHSHSYAQMVRVLRRTASRCARHARTYSGRFDGRELIVIEFSSRPGQHLMPEVK 231  
DB 8 ISFEYHRYPELRREALVSVMLQCTAISRIYTVGRSFEGRLLVIELSDNPVGHEPEFK 67  
QY 232 LIGNIHGNEVAGREMLIYLAQYLCEYLLGNPRIQRLNTRIHLLPSINPDGYEVA 291  
DB 68 YIGMHGNEAVGRELLIFLAQYLCNEYQKGNETIVNLHSTRIHIMPSLNPDGFEKASQ 127  
QY 292 GAGYNGWTSQRQNAQNLNRPDLTSEYVRLAETGARS DHI-----PIQHYWVGK 345  
DB 128 PGLKQWVGRSNAQGLDLNRPDLDRIVY-VNEKEGGNNHLLKNMKIVDQN---TK 183  
QY 346 VAPETKAIMKWMQTIPIFVLSASLHGGDLVSYPPDFSKHPQEEKMFSPPTDEKMFLLSR 405  
DB 184 LAPETKAVIHWINDIPFVLSANLHGGDLVANYPYDETSGSAHE-YSSSPDDAIFQSLAR 242  
QY 406 AYADVHPMMDRSENRCGN-----FLKRGSIINGADWYSFTGMSDFNYLHNCFEITV 460  
DB 243 AYSSFNPMSPDNPFPCKRKNDDSSFV--DGTNGGAWYSVPGMQDFNYLSSNCFEITV 300  
QY 461 ELGCVKFPPEEALYTLQHNKESLLNFVETVHRGKGVVTDKFGKPKVNARI SVKGRHD 520  
DB 301 ELSCEKFPPEETLKTWEDNKNLSIYLEQIHGRGVKGFVRDLQGNPIANATISVEGIDHD 360  
QY 521 ITTAPDGDYWRLLPPGHIHIVIAQAPGKAVIKKVIIPARKMRAGRVDFILO 571  
DB 361 VTSKAGDGYWRLLIPGNYKLTASAPGYLATKKVAVP--YSPAAGVDFELE 409

RESULT 10  
PCT-US96-07528-2  
Sequence 2, Application PC/TUS9607528  
GENERAL INFORMATION:  
APPLICANT: The Jackson Laboratory  
APPLICANT: Naggett, Jurgen K.  
TITLE OF INVENTION: CLINICAL DISORDERS ASSOCIATED WITH  
TITLE OF INVENTION: CARBOXYPEPTIDASE E MUTATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kevin M. Farrell, P.C.  
STREET: P.O. Box 999  
CITY: York Harbor

STATE: ME  
COUNTRY: US  
ZIP: 03911  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/07528  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/452,262  
FILING DATE: 26-MAY-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Farrell, Kevin M.  
REGISTRATION NUMBER: 35,505  
REFERENCE/DOCKET NUMBER: JL-9501 WO  
TELEPHONE: (207) 363-0558  
TELEFAX: (207) 363-0528  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US96-07528-2

Query Match 30.2%; Score 1051; DB 5; Length 435;

Best Local Similarity 50.1%; Pred. No. 2.3e-96; Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;

QY 172 IRFSHSHYAQMVRVLRRTASRCARHARTYISGRSFDGRELIVIEFSSRPGQHELMPEYK 231  
DB 8 ISPEYHRYPELREALVSVMLQCTAISRIYTVGRSPFEGRELLVIELSDNPGVHEPGEPEK 67  
QY 232 LGNIHGNVAGREMLIYLAQVLCSEYLLGNPRIQLNTTRIHLPSINPDGYEVAAE 291  
DB 68 YIGNMHGNEAVGRELLIFLAQVLCNEYQGNETIVNLHSTRIHIMPSLNPDGFEKAAQ 127  
QY 292 GAGYNGWTSGRONALNDRNFPDLTSEYRLAETRGARSDHI-----PIQHYWGWK 345  
DB 128 PGELKDFWGRSNAQGLDNRNFPDLDRIVY-VNEKEGGPNHLLKMKKIVDQN---TK 183  
QY 346 VAPETKAIMKMQTTPFVLSASLHGGDLVVSYPDFSKHPQBEKMFSPPTDEKMKLLSR 405  
DB 184 LAPETKAVIHWIMDIPFVLSANLHGGDLVANYPYDETRSGSAHE-YSSSPDDAIFQSLAR 242  
QY 406 AYADVHPMMDRSENRCGN-----FLKRGSIINGADWYSFTGGMSDFNYLHTNCFEIV 460  
DB 243 AYSSFPNMSDPNRPCKRKNDDSSFV--DGTNTGGAWYSVPGQMDFNYLSSNCFEIV 300  
QY 461 ELGCVKFPPEEALYTLWQHNKESLLNFVETVHRGKIGVVTDKFGKPKVKNARISVKGIRHD 520  
DB 301 ELSCFKFPPEETLKTWEDKNLSLIEQIHRGVKGFVRDLQGNPIANATISVGDIDH 360  
QY 521 ITTAPDGDYWRLLPPGHIHIVIAQAGYAKVKKVILPARKMGAGRVDFILQ 571  
DB 361 VTSKADGDYWRLLIPGNYKLTASAPGYLAITKKVAVP--YSPAAGVDVFELE 409

RESULT 11

US-08-111-939-14  
Sequence 14, Application US/08111939  
Patent No. 5460951  
GENERAL INFORMATION:  
APPLICANT: Kawai, Shinji  
APPLICANT: Takeshita, Sunao  
APPLICANT: Okazaki, Makoto  
APPLICANT: Amann, Egon

TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like  
TITLE OF INVENTION: Protein and Process for its Production  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/111,939  
FILING DATE: 26-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 324033/92  
FILING DATE: 03-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 230029/92  
FILING DATE: 28-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Forman, David S.  
REGISTRATION NUMBER: 33,694  
REFERENCE/DOCKET NUMBER: 02481.1321-00000  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4000  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-111-939-14

Query Match 30.2%; Score 1050; DB 1; Length 435;

Best Local Similarity 50.1%; Pred. No. 2.9e-96;

Matches 206; Conservative 65; Mismatches 120; Indels 20; Gaps 7;

QY 172 IRFSHSHYAQMVRVLRRTASRCARHARTYISGRSFDGRELIVIEFSSRPGQHELMPEYK 231  
DB 8 ISPEYHRYPELREALVSVMLQCTAISRIYTVGRSPFEGRELLVIELSDNPGVHEPGEPEK 67  
QY 232 LGNIHGNVAGREMLIYLAQVLCSEYLLGNPRIQLNTTRIHLPSINPDGYEVAAE 291  
DB 68 YIGNMHGNEAVGRELLIFLAQVLCNEYQGNETIVNLHSTRIHIMPSLNPDGFEKAAQ 127  
QY 292 GAGYNGWTSGRONALNDRNFPDLTSEYRLAETRGARSDHI-----PIQHYWGWK 345  
DB 128 PGELKDFWGRSNAQGLDNRNFPDLDRIVY-VNEKEGGPNHLLKMKKIVDQN---SK 183  
QY 346 VAPETKAIMKMQTTPFVLSASLHGGDLVVSYPDFSKHPQBEKMFSPPTDEKMKLLSR 405  
DB 184 LAPETKAVIHWIMDIPFVLSANLHGGDLVANYPYDETRSGTAHE-YSSCPDDAIFQSLAR 242  
QY 406 AYADVHPMMDRSENRCGN-----FLKRGSIINGADWYSFTGGMSDFNYLHTNCFEIV 460  
DB 243 AYSSFPNMSDPNRPCKRKNDDSSFV--DGTNTGGAWYSVPGQMDFNYLSSNCFEIV 300  
QY 461 ELGCVKFPPEEALYTLWQHNKESLLNFVETVHRGKIGVVTDKFGKPKVKNARISVKGIRHD 520  
DB 301 ELSCFKFPPEETLKSXYWEDKNLSLYLEQIHRGVKGFVRDLQGNPIANATISVGDIDH 360  
QY 521 ITTAPDGDYWRLLPPGHIHIVIAQAGYAKVKKVILPARKMGAGRVDFILQ 571  
DB 361 VTSKADGDYWRLLIPGNYKLTASAPGYLAITKKVAVP--FSPAAGVDVFELE 409

## RESULT 12

US-09-233-989-6  
; Sequence 6, Application US/09233989  
; Patent No. 6248527  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Meyer, Joanne  
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on  
; TITLE OF INVENTION: Mutations Found in Carboxypeptidase E  
; FILE REFERENCE: 5800-14, 035800/174130  
; CURRENT APPLICATION NUMBER: US/09/233,989  
; CURRENT FILING DATE: 1999-01-19  
; EARLIER APPLICATION NUMBER: 60/105,102  
; EARLIER FILING DATE: 1998-10-21  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 476  
; TYPE: PRT  
; ORGANISM: Murinae gen. sp.  
; FEATURE:  
; OTHER INFORMATION: carboxypeptidase E  
US-09-233-989-6

Query Match 30.1%; Score 1049; DB 3; Length 476;  
Best Local Similarity 48.0%; Pred. No. 4.3e-96;  
Matches 211; Conservative 67; Mismatches 130; Indels 32; Gaps 8;  
  
Qy 155 GGLADALPSGLPPTIR-----FSHSYAQWVRVLRRTASCAHVARTYSI 202  
Db 20 GLLTAAEQEGAPAGARRRRRLQEDGISFEYHRYPELRREALVSVMLQCTAISRYTV 79  
  
Qy 203 GRSDGRELIVIESSPGOHELMPEVKLIGNTHGNEVAGREMLIYLAQYLCEYLLGN 262  
Db 80 GRSPGRELIVIELSDNPNVGHPEPEPKYIGNMHGNEAVGRELLIFLAQYLCEYKGN 139  
  
Qy 263 PRIQLLNTTRIHLLPSINPDGYAAAGAGYNGWTSGRONQNLNLRNFPDLTSEY 322  
Db 140 ETIVNLHSTRIHIMPSLNPDPGEKAASQPELKDWFVGRSNAQIDLNRFDPDLRIYV 199  
  
Qy 323 RLATRGARSDHI-----PIPOHYWGWKVAPEYKAIKMWMTTIPFVLSASLHGGDLVVS 376  
Db 200 -VNEKEGGPNHLLKLNKKIVDQN---SKLAPETKAVIHWIMDIPFVLSANLHGGDLVAN 255  
  
Qy 377 YPDPFKHPQEEKMFSPTPEKMKLISRAYADVHPMMDESENRCGN-----FLKRG 431  
Db 256 YPYDETRSGTAHE-YSSCPDDAIFQSLARAYSSFNPMVSDPNRPPCKNDSDSFV--DG 312  
  
Qy 432 IINGADWYSFTGGMDSFNHLHNTCFEITVELGCVKFPPEALYTLQHNKESLLNFVETV 491  
Db 313 TTNGGANYSPGGMDFNYSLSNCFEITVELSCFPPEETLKSYWEDNKNLSISYLEQI 372  
  
Qy 492 HRGIGVVVTKFGKPVKNARISVKGRHDIITAPDGYWRLPFGIHIYIAQAFYARVI 551  
Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGWYRLAPGNKLTASAPGYLAIT 432  
  
Qy 552 KKVIIIPARMKAGRVDFILO 571  
Db 433 KKVAVP--FSPAVGVDFELE 450

## RESULT 13

US-08-111-939-15  
; Sequence 15, Application US/08111939  
; Patent No. 5460951  
; GENERAL INFORMATION:  
; APPLICANT: Kawai, Shinji  
; APPLICANT: Takeshita, Sunao  
; APPLICANT: Okazaki, Makoto  
; APPLICANT: Amann, Egon  
; TITLE OF INVENTION: Bone-Related Carboxypeptidase-Like  
; TITLE OF INVENTION: Protein and Process for its Production

NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/111,939  
FILING DATE: 26-AUG-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 324033/92  
FILING DATE: 03-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 230029/92  
FILING DATE: 28-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Foxman, David S.  
REGISTRATION NUMBER: 33,694  
REFERENCE/DOCKET NUMBER: 02481.1321-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4000  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 435 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-111-939-15

Query Match 30.1%; Score 1047; DB 1; Length 435;  
Best Local Similarity 49.9%; Pred. No. 5.8e-96;  
Matches 205; Conservative 66; Mismatches 120; Indels 20; Gaps 7;  
  
Qy 172 IRFSHSYAQWVRVLRRTASCAHVARTYSIGRSFDGRELIVIEFSSRPGOHELMPEVK 231  
Db 8 ISFEYHRYPELRREALVSVMLQCTAISRYTVGTFEGRELLVIELSDNPGVHPEGEPEFK 67  
  
Qy 232 LIGNHNEVAGREMLIYLAQYLCEYLLGNPRIQLLNTTRIHLLPSINPDGYEVAAS 291  
Db 68 YIGNMHGNEAVGRELLIFLAQYLCEYQNGNETIVNLHSTRIHIMPSLNPDPGEKAASQ 127  
  
Qy 292 GAGYNGWTSGRONQNLNLRNFPDLTSEYRLAETRGARSDHI-----PIPOHYWGWK 345  
Db 128 PGLKDFVGRSNAQIDLNRFDPDLRIYV-VNEKEGGPNHLLKLNKKIVDQN---SK 183  
  
Qy 346 VAPETKAIKMWMTTIPFVLSASLHGGDLVVSYPDFSKHPQEEKMFSPTPEKMKFLLSR 405  
Db 184 LAPETKAVIHWIMDIPFVLSANLHGGDLVANYPDETRSGTAHE-YSSCPDDAIFQSLAR 242  
  
Qy 406 AYADVHPMMDRSENRCGN-----FLKRGSIINGADWYSFTGGMDSFNHLHNTCFEITV 460  
Db 243 AYSSFNPMVSDPNRPPCKNDSDSFV--DGTNGGAWYSVPGGMQDFNYSLSNCFEITV 300  
  
Qy 461 ELGCVKFPPEALYTLQHNKESLLNFVETVHRGIGVVVTKFGKPVKNARISVKGRH 520  
Db 301 ELSCFPPEETLKSYWEDNKNLSINYLEQIHRGVKGFVRDLQGNPIANATISVDGIDHD 360  
  
Qy 521 ITTAPDGYWRLPFGIHIYIAQAFYARVIKVKVIIIPARMKAGRVDFILO 571  
Db 361 VTSKADGDYWRLPFGNPKLTASAPGYLAITKKVAVP--FSPAVGVDFELE 409  
  
RESULT 14



US-09-641-741-2  
; Sequence 2, Application US/09641741  
; Patent No. 6420155  
; GENERAL INFORMATION:  
; APPLICANT: Kerry E. Quinn  
; TITLE OF INVENTION: Aortic Carboxypeptidase-Like Proteins and Nucleic Acids  
; FILE REFERENCE: 15966-581  
; CURRENT APPLICATION NUMBER: US/09/641,741  
; EARLIER FILING DATE: 2000-08-18  
; PRIOR APPLICATION NUMBER: 60/159,613  
; PRIOR FILING DATE: 1999-10-14  
; PRIOR APPLICATION NUMBER: 60/175,534  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/224,086  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 734  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-641-741-2

Query Match 29.8%; Score 1036; DB 4; Length 734;  
Best Local Similarity 42.5%; Pred. No. 1.8e-94;  
Matches 234; Conservative 81; Mismatches 161; Indels 74; Gaps 17;

QY 73 VLHQLLEGQCNPLRLGCAVLAPRCGGWVRPCRHI CEGLRE--VCOPAFDAIDMAMP 130  
DB 239 VLNLLPEPQVAFIRLLPQTLW----QGG--APC-----LRAEILACFVSDPNDL--- 282  
QY 131 YFLDCHRYFTREDEGCYDPLEKRGLEADEALPSGLPTTFIRPSHSVAQWVRVLRRTA 190  
DB 283 -FLEA-----PASGSSDPLD-----FOHNYKAMKRLMKQVQ 313  
QY 191 SRCACHVARTYSIGRSFDGRELIVIEFSRPGQHELMPEVKLIGNTHGNEVAGREMLIYL 250  
DB 314 EQCPNTRIYSIGSKSQGLKLYWMSDKGEHELGEPEVRYVAGHNEALGRELLILL 373  
QY 251 AQYLCSEYLLGNPRIORLLNTTTRIHLPLSINPDGYEVAAGAGYNGWTSGRQNAQNLDL 310  
DB 374 MQFLCHEFLKGNPRVTRLLSEMRHLLPSNPDGYEIAVHRGSELVGAEGRWNNQSIDL 433  
QY 311 NRNFPDLTSEYYRLATRG-----ARSDHIPQHYTW--GKVAPETKAIMKWMQTIPIFY 363  
DB 434 NHNFADLNTPLWB-AQDDGKVPHIVPNHLLPLTYTTLPNATVAPETRAVIKWKRIPIFY 492  
QY 364 LSASLHGGLLVSPYDFDSKHPQEEKMFSPPTDEKMFKLLSRAYADVHPMMMDRSNRC- 422  
DB 493 LSNALHGGLLVSPYDFDMTRTPWAARELPTPDADVFRMLSTVYAGSNLAMOQTSRRPCH 552  
QY 423 GGNFLKRGSIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWOHNKE 482  
DB 553 SQDFSVDHNIINGADWHTVPGSMNDSFYLHTNCFEITVELSCDKFPHENELPQEWENKD 612  
QY 483 SLINFVETVHRGKGVVTKFGK- PVKNARI SVKGI RHDITTA PDGDYWRLLPPGHIHVI 541  
DB 613 ALLTYLEQVRMGVAVVRDKDTBELGIADAVIADVGINHDVTTAWGDDYWRLLTPGDYMT 672  
QY 542 AQAPGVAKVKKVIIIPARMKACRGRVDFILQPLGMGPKNFTHGIRRTGRPHDPLGGASLGE 601  
DB 673 ASAEGHVSTRNC-----RVTFEGFP---PCNFV---LTKT-PKQRLRELLAAGA 716  
QY 602 ATEPDPFLRAR 611  
DB 717 KVPPD-LRRR 725

RESULT 15  
US-09-233-989-4  
; Sequence 4, Application US/09233989

; Patent No. 6248527  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Meyer, Joanne  
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on  
; FILE REFERENCE: 5800-14, 035800/174130  
; CURRENT APPLICATION NUMBER: US/09/233,989  
; CURRENT FILING DATE: 1999-01-19  
; EARLIER APPLICATION NUMBER: 60/105,102  
; EARLIER FILING DATE: 1998-10-21  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 454  
; TYPE: PRT  
; ORGANISM: lopam  
; FEATURE:  
; OTHER INFORMATION: carboxypeptidase E  
US-09-233-989-4

Query Match 29.7%; Score 1035; DB 3; Length 454;  
Best Local Similarity 50.1%; Pred. No. 1e-94;  
Matches 207; Conservative 65; Mismatches 121; Indels 20; Gaps 7;

QY 172 IRFSSHSYAQMVRVLRRTASRCACHVARTYSIGRSFDGRELIVIEFSRPGQHELMPEVK 231  
DB 27 ISFHYRHEELRKALVSVWLQCTTIARIYITIGSEFGEELLVLEMSDNGTHEPEPEPK 86  
QY 232 LTGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIORLLNTTTRIHLPLSINPDGYEVAAG 291  
DB 87 YIANMHGNEAVGRELLIYLAQYLCNQYQOGNETIIDLIHSTRIHLMPSNPDGFEKAAQ 146  
QY 292 GAGYNGWTSGRQNAQNLDLNRNFPDLTSEYYRLAETRGARSDHI-----PIPOHYWGWK 345  
DB 147 PGEIKDWFVGRSNAQGVDLNRNFPDLDRIIY-TNEREGGANNHLLQNKKAVDEN---TK 202  
QY 346 VAPETKAIMKWMQTIPIFVLSASLHGGLLVSPYDFDSKHPQEEKMFSPPTDEKMFKLLSR 405  
DB 203 LAPETKAVIHWIMEIPFVLSANLHGGLVAVVANTPYDETITRGSTHE-YSASPDVIVFKSLAK 261  
QY 406 AYADVHPMMMDRSNRC-----GGNFKRGSIIINGADWYSFTGMSDFNYLHTNCFEITV 460  
DB 262 AFSIYNPVMSDPQRPCKRKHDDSSF--KDGITNGGAWYSVPGMQDFNYLSSNCFEITL 319  
QY 461 ELGCVKFPPEEALYTLWOHNKESLNFVETVHRGKGVVTKFGKVPVKNARISVKGI RHD 520  
DB 320 ELSCDKFFNEEDLTQTYWEQNRNSLVNYIEQVHRGVKGVYRDLQGNPIFNATISVEGIDHD 379  
QY 521 ITTAPGDYWRLLPPGIHIVIAQAPGVAKVKKVIIIPARMKACRGRVDFILQPL 573  
DB 380 ITTAKDGDYWRLLRROGNKVAASAPGYLTUVIKKVAVP--HSPATRVDFPELES 430

Search completed: January 25, 2005, 08:53:56  
Job time : 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2005, 08:52:30 ; Search time 149 Seconds  
(without alignments)  
1554.272 Million cell updates/sec

Title: US-10-757-262-104

Perfect score: 3480

Sequence: 1 MPPPPLLLLLVVAAARP.....MWSYFTSLTHRPRLKY 641

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3480	100.0	641	9	US-09-827-040-10
2	3480	100.0	641	17	US-10-757-262-104
3	3480	100.0	641	17	US-10-723-860-914
4	1849.5	53.1	364	14	US-10-094-749-2481
5	1058	30.4	476	13	US-10-060-425-12
6	1057	30.4	476	9	US-09-827-040-2
7	1054	30.3	476	9	US-09-827-040-2
8	1054	30.3	476	13	US-10-060-425-4
9	1054	30.3	476	14	US-10-295-027-576
10	1054	30.3	476	17	US-10-723-860-720
11	1052	30.2	434	9	US-10-827-040-7
12	1052	30.2	434	13	US-10-060-425-15
13	1051	30.2	491	9	US-09-925-300-1715

14	1051	30.2	491	11	US-09-833-245-1035	Sequence 1035, Ap
15	1051	30.2	491	11	US-09-833-245-1057	Sequence 1057, Ap
16	1049	30.1	476	9	US-09-827-040-6	Sequence 6, Appli
17	1049	30.1	476	13	US-10-060-425-14	Sequence 14, Appl
18	1037	29.8	734	14	US-10-290-078-24	Sequence 24, Appl
19	1036	29.8	714	15	US-10-307-817-158	Sequence 158, App
20	1036	29.8	720	15	US-10-307-817-162	Sequence 162, App
21	1036	29.8	734	10	US-09-996-015-2	Sequence 2, Appli
22	1036	29.8	734	13	US-10-052-586-458	Sequence 458, App
23	1036	29.8	734	14	US-10-174-590-458	Sequence 458, App
24	1036	29.8	734	14	US-10-176-758-458	Sequence 458, App
25	1036	29.8	734	14	US-10-175-737-458	Sequence 458, App
26	1036	29.8	734	14	US-10-174-581-458	Sequence 458, App
27	1036	29.8	734	14	US-10-176-483-458	Sequence 458, App
28	1036	29.8	734	14	US-10-176-749-458	Sequence 458, App
29	1036	29.8	734	14	US-10-176-914-458	Sequence 458, App
30	1036	29.8	734	14	US-10-176-915-458	Sequence 458, App
31	1036	29.8	734	14	US-10-173-706-458	Sequence 458, App
32	1036	29.8	734	14	US-10-175-738-458	Sequence 458, App
33	1036	29.8	734	14	US-10-175-752-458	Sequence 458, App
34	1036	29.8	734	14	US-10-176-482-458	Sequence 458, App
35	1036	29.8	734	14	US-10-176-757-458	Sequence 458, App
36	1036	29.8	734	14	US-10-176-913-458	Sequence 458, App
37	1036	29.8	734	14	US-10-180-552-458	Sequence 458, App
38	1036	29.8	734	14	US-10-180-557-458	Sequence 458, App
39	1036	29.8	734	14	US-10-173-700-458	Sequence 458, App
40	1036	29.8	734	14	US-10-174-572-458	Sequence 458, App
41	1036	29.8	734	14	US-10-174-579-458	Sequence 458, App
42	1036	29.8	734	14	US-10-174-582-458	Sequence 458, App
43	1036	29.8	734	14	US-10-174-588-458	Sequence 458, App
44	1036	29.8	734	14	US-10-175-739-458	Sequence 458, App
45	1036	29.8	734	14	US-10-175-740-458	Sequence 458, App

ALIGNMENTS

RESULT 1

US-09-827-040-10  
; Sequence 10, Application US/09827040  
; Patent No. US20010024792A1  
; GENERAL INFORMATION:  
; APPLICANT: Meyer, Joanne  
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on  
; TITLE OF INVENTION: Mutations Found in Carboxypeptidase E  
; FILE REFERENCE: 5800-14, 035800/174130  
; CURRENT APPLICATION NUMBER: US/09/827,040  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: 09/233,989  
; PRIOR FILING DATE: 1999-01-19  
; PRIOR APPLICATION NUMBER: 60/105,102  
; PRIOR FILING DATE: 1998-10-21  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 641  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: carboxypeptidase homolog -- CPZ  
US-09-827-040-10  
Query Match 100.0%; Score 3480; DB 9; Length 641;  
Best Local Similarity 100.0%; Pred. No. 2.3e-315;  
Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MPPPPLLLLLVVAAARPCGFERNPAATCVDLQRTCSDAAYNHTTFFNLQHSWE 60  
Db 1 MPPPPLLLLLVVAAARPCGFERNPAATCVDLQRTCSDAAYNHTTFFNLQHSWE 60  
Qy 61 VFASSEYIILSVLHQLLEGCCNPDRLGLGCAVLAPRCGGWVRPCHCEGLREVCQP 120

Db 61 VVEASSEYILLSVLHQLLEGQCNPDRLRLGCAVLAPRCGCGWVRPRCRHICEGLREVCP 120  
Qy 121 AFDAIDMWPFLDCHRYFTREDEGCYDPLEKLRGLEADEALPSGLPPTFIRFHHSYA 180  
Db 121 AFDAIDMWPFLDCHRYFTREDEGCYDPLEKLRGLEADEALPSGLPPTFIRFHHSYA 180  
Qy 181 QMVRVLRRTASCAHVARTYSIGRSPDGRRELLVIEFSSRPGQHELMPEVKLIGNHNE 240  
Db 181 QMVRVLRRTASCAHVARTYSIGRSPDGRRELLVIEFSSRPGQHELMPEVKLIGNHNE 240  
Qy 241 VAGREMLIYLAQVLCSEYLLGNPRIQRLNLTTRIHLPSINPDGYEVAAGAGYNGWTS 300  
Db 241 VAGREMLIYLAQVLCSEYLLGNPRIQRLNLTTRIHLPSINPDGYEVAAGAGYNGWTS 300  
Qy 301 GRQNAQNLNPNFPDLTSEYRLAETRGARSDHIPPOHYWGWKVAPETKAIKMWQTI 360  
Db 301 GRQNAQNLNPNFPDLTSEYRLAETRGARSDHIPPOHYWGWKVAPETKAIKMWQTI 360  
Qy 361 PFVLSASLHGGDLVSYPPDFSKHPQEEKMFSPPTDEKMFKLLSRAYADVHPMMDRSEN 420  
Db 361 PFVLSASLHGGDLVSYPPDFSKHPQEEKMFSPPTDEKMFKLLSRAYADVHPMMDRSEN 420  
Qy 421 RCGNFKRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGCVKFPPEEALYTLWQHN 480  
Db 421 RCGNFKRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGCVKFPPEEALYTLWQHN 480  
Qy 481 KESLLNFVETVHRGKGVVTDKFGPKVNARI SVKGI RHDITTA PDGDYWRLLPFGIHI V 540  
Db 481 KESLLNFVETVHRGKGVVTDKFGPKVNARI SVKGI RHDITTA PDGDYWRLLPFGIHI V 540  
Qy 541 IQAAGYAKVKKVVIIPARMKAGRVDFILOPLGMPKPNFIHGLRRTGPHDPLGGASSLG 600  
Db 541 IQAAGYAKVKKVVIIPARMKAGRVDFILOPLGMPKPNFIHGLRRTGPHDPLGGASSLG 600  
Qy 601 EATEPDLRARRQPSADGSKPMWWSYFTSLSTRPRWLLKY 641  
Db 601 EATEPDLRARRQPSADGSKPMWWSYFTSLSTRPRWLLKY 641

RESULT 2  
US-10-757-262-104  
; Sequence 104, Application US/10757262  
; Publication No. US20040197825A1  
; GENERAL INFORMATION:  
; APPLICANT: Karicheti, Venkateswarlu  
; APPLICANT: Silas-Santiago, Immaculada  
; APPLICANT: Eliasof, Scott D.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING  
; TITLE OF INVENTION: UROLOGICAL DISORDERS USING 44390, 54181, 211, 5687, 884,  
; TITLE OF INVENTION: 1405, 636, 4421, 5410, 30905, 2045, 16405, 18560, 2047,  
; TITLE OF INVENTION: 33751, 52872, 14063, 20739, 32544, 43239, 44373, 51164,  
; TITLE OF INVENTION: 53010, 16852, 1587, 2207, 22445, 2387, 52908, 69112, 14990,  
; TITLE OF INVENTION: 18547, 115, 579, 15985, 15625, 760, 18603, 2395, 2554, 8675,  
; TITLE OF INVENTION: 32720, 4809, 14303, 16816, 17827, 32620, 577, 619, 1423,  
; TITLE OF INVENTION: 2158, 8263, 15402, 16209, 16386, 21165, 30911, 41897, 1643,  
; TITLE OF INVENTION: 2543, 9626, 13231, 32409, 84260, 2882, 8203, 32678 OR  
; FILE REFERENCE: MPI03-007PIRNONMIM  
; CURRENT APPLICATION NUMBER: US/10757,262  
; CURRENT FILING DATE: 2004-01-14  
; PRIOR APPLICATION NUMBER: US 60/440,318  
; PRIOR FILING DATE: 2003-01-15  
; PRIOR APPLICATION NUMBER: US 60/444,783  
; PRIOR FILING DATE: 2003-02-04  
; PRIOR APPLICATION NUMBER: US 60/457,901  
; PRIOR FILING DATE: 2003-03-27  
; PRIOR APPLICATION NUMBER: US 60/468,775  
; PRIOR FILING DATE: 2003-05-08  
; PRIOR APPLICATION NUMBER: US 60/471,614  
; PRIOR FILING DATE: 2003-05-19  
; PRIOR APPLICATION NUMBER: US 60/478,742  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: US 60/488,529

; PRIOR FILING DATE: 2003-07-18  
; PRIOR APPLICATION NUMBER: US 60/491,156  
; PRIOR FILING DATE: 2003-07-30  
; PRIOR APPLICATION NUMBER: US 60/499,594  
; PRIOR FILING DATE: 2003-09-02  
; PRIOR APPLICATION NUMBER: US 60/506,332  
; PRIOR FILING DATE: 2003-09-26  
; NUMBER OF SEQ ID NOS: 136  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 104  
; LENGTH: 641  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-757-262-104

Query Match 100.0%; Score 3480; DB 17; Length 641;  
Best Local Similarity 100.0%; Pred. No. 2.3e-315;  
Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPPPLLLLLLTVLVAAARPGCEPERNPAATCVDLQLRTCSDAAYNHTTFFPNLLQHRWE 60  
Db 1 MPPPLLLLLLTVLVAAARPGCEPERNPAATCVDLQLRTCSDAAYNHTTFFPNLLQHRWE 60  
Qy 61 VVEASSEYILLSVLHQLLEGQCNPDRLRLGCAVLAPRCGCGWVRPRCRHICEGLREVCP 120  
Db 61 VVEASSEYILLSVLHQLLEGQCNPDRLRLGCAVLAPRCGCGWVRPRCRHICEGLREVCP 120  
Qy 121 AFDAIDMWPFLDCHRYFTREDEGCYDPLEKLRGLEADEALPSGLPPTFIRFHHSYA 180  
Db 121 AFDAIDMWPFLDCHRYFTREDEGCYDPLEKLRGLEADEALPSGLPPTFIRFHHSYA 180  
Qy 181 QMVRVLRRTASCAHVARTYSIGRSPDGRRELLVIEFSSRPGQHELMPEVKLIGNHNE 240  
Db 181 QMVRVLRRTASCAHVARTYSIGRSPDGRRELLVIEFSSRPGQHELMPEVKLIGNHNE 240  
Qy 241 VAGREMLIYLAQVLCSEYLLGNPRIQRLNLTTRIHLPSINPDGYEVAAGAGYNGWTS 300  
Db 241 VAGREMLIYLAQVLCSEYLLGNPRIQRLNLTTRIHLPSINPDGYEVAAGAGYNGWTS 300  
Qy 301 GRQNAQNLNPNFPDLTSEYRLAETRGARSDHIPPOHYWGWKVAPETKAIKMWQTI 360  
Db 301 GRQNAQNLNPNFPDLTSEYRLAETRGARSDHIPPOHYWGWKVAPETKAIKMWQTI 360  
Qy 361 PFVLSASLHGGDLVSYPPDFSKHPQEEKMFSPPTDEKMFKLLSRAYADVHPMMDRSEN 420  
Db 361 PFVLSASLHGGDLVSYPPDFSKHPQEEKMFSPPTDEKMFKLLSRAYADVHPMMDRSEN 420  
Qy 421 RCGNFKRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGCVKFPPEEALYTLWQHN 480  
Db 421 RCGNFKRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGCVKFPPEEALYTLWQHN 480  
Qy 481 KESLLNFVETVHRGKGVVTDKFGPKVNARI SVKGI RHDITTA PDGDYWRLLPFGIHI V 540  
Db 481 KESLLNFVETVHRGKGVVTDKFGPKVNARI SVKGI RHDITTA PDGDYWRLLPFGIHI V 540  
Qy 541 IQAAGYAKVKKVVIIPARMKAGRVDFILOPLGMPKPNFIHGLRRTGPHDPLGGASSLG 600  
Db 541 IQAAGYAKVKKVVIIPARMKAGRVDFILOPLGMPKPNFIHGLRRTGPHDPLGGASSLG 600  
Qy 601 EATEPDLRARRQPSADGSKPMWWSYFTSLSTRPRWLLKY 641  
Db 601 EATEPDLRARRQPSADGSKPMWWSYFTSLSTRPRWLLKY 641

RESULT 3  
US-10-723-860-914  
; Sequence 914, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &

```
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 914
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-914

Query Match      100.0%; Score 3480; DB 17; Length 641;
Best Local Similarity 100.0%; Pred. No. 2.3e-315;
Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPPPPLLLLLVVAAARPGCEPERNPAATCDVQLQRTCSDAAYNHTTFFNLLQHSWE 60
Db 1 MPPPPLLLLLVVAAARPGCEPERNPAATCDVQLQRTCSDAAYNHTTFFNLLQHSWE 60

Qy 61 VVEASSYILLVLLHQLLEGOCNPDRLILGCACVLAAPCEGQWVRPCRHICEGLREVCOP 120
Db 61 VVEASSYILLVLLHQLLEGOCNPDRLILGCACVLAAPCEGQWVRPCRHICEGLREVCOP 120

Qy 121 AFAIDAMWPFYLDCHRYFTREDEGCYDPLEKLGGLLEADEALPSGLPPTFIHFSSHSYA 180
Db 121 AFAIDAMWPFYLDCHRYFTREDEGCYDPLEKLGGLLEADEALPSGLPPTFIHFSSHSYA 180

Qy 181 QMVRVLRRTASCAHVAARTYISGRSFDGRELIVIEFSSRPGQHELMPEVKLIGNIHNE 240
Db 181 QMVRVLRRTASCAHVAARTYISGRSFDGRELIVIEFSSRPGQHELMPEVKLIGNIHNE 240

Qy 241 VAGREMLIYLAOYLCSYLLGNPRIORLLNTTTRIHLPSINPDGYEVAAAEAGAGYNGWTS 300
Db 241 VAGREMLIYLAOYLCSYLLGNPRIORLLNTTTRIHLPSINPDGYEVAAAEAGAGYNGWTS 300

Qy 301 GRONAQNLDNRNPDDLTSEYRLAETRGARSDHIPIPOHYWGWKVAPETKAIKWMQTI 360
Db 301 GRONAQNLDNRNPDDLTSEYRLAETRGARSDHIPIPOHYWGWKVAPETKAIKWMQTI 360

Qy 361 PFVLSASLHGDLVSVYFPDFSKHPQBEKMFSPPTDEKMPKLLSRAYADVHPMMDRSEN 420
Db 361 PFVLSASLHGDLVSVYFPDFSKHPQBEKMFSPPTDEKMPKLLSRAYADVHPMMDRSEN 420

Qy 421 RCGNFIKRSIINGADWYFTGMSDFNVLHNTCFEITVELGCVKFPPEALYTLQHN 480
Db 421 RCGNFIKRSIINGADWYFTGMSDFNVLHNTCFEITVELGCVKFPPEALYTLQHN 480

Qy 481 KESLLNFVETVHRGIKGVWTDKFGKPKVNARISVKGIRHDITAPDGDYWRLLPPGHIIV 540
Db 481 KESLLNFVETVHRGIKGVWTDKFGKPKVNARISVKGIRHDITAPDGDYWRLLPPGHIIV 540

Qy 541 IAQAPGVAKVKKVIIIPARMKRAGRVDFILQPLGMGPKNFHGLRRTGPHDPLGGASSLG 600
Db 541 IAQAPGVAKVKKVIIIPARMKRAGRVDFILQPLGMGPKNFHGLRRTGPHDPLGGASSLG 600

Qy 601 EATEPDLRARRQPSADGSPKPMWSYFTSLSTRPRWLLKY 641
Db 601 EATEPDLRARRQPSADGSPKPMWSYFTSLSTRPRWLLKY 641

RESULT 4
US-10-094-749-2481
; Sequence 2481, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
```

```
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2481
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2481

Query Match      53.1%; Score 1849.5; DB 14; Length 364;
Best Local Similarity 96.4%; Pred. No. 1.6e-163;
Matches 344; Conservative 1; Mismatches 1; Indels 11; Gaps 1;

Qy 1 MPPPPLLLLLVVAAARPGCEPERNPA-----ATCVDLQRLTCSDAAYNHTT 49
Db 1 MPPPPPPLLLLLVVAAARPGCEPERNPAECHRPPAADSATCVDLQRLTCSDAAYNHTT 60

Qy 50 FPNLLQHSWEVVEASSEYILLVLLHQLLEGOCNPDRLILGCACVLAAPCEGQWVRPCRH 109
Db 61 FPNLLQHSWEVVEASSEYILLVLLHQLLEGOCNPDRLILGCACVLAAPCEGQWVRPCRH 120

Qy 110 ICSEGLREVCQAPFADIDMAMPYFLDCHRYFTREDEGCYDPLEKLGGLLEADEALPSGLPP 169
Db 121 ICSEGLREVCQAPFADIDMAMPYFLDCHRYFTREDEGCYDPLEKLGGLLEADEALPSGLPP 180

Qy 170 TFIHFSSHSYAQMVRVLRRTASCAHVAARTYISGRSFDGRELIVIEFSSRPGQHELMPE 229
Db 181 TFIHFSSHSYAQMVRVLRRTASCAHVAARTYISGRSFDGRELIVIEFSSRPGQHELMPE 240

Qy 230 VKLIGNIHGNEVAGREMLIYLAOYLCSYLLGNPRIORLLNTTTRIHLPSINPDGYEVAA 289
Db 241 VKLIGNIHGNEVAGREMLIYLAOYLCSYLLGNPRIORLLNTTTRIHLPSINPDGYEVAA 300

Qy 290 AEGAGYNGWTSGRQAQNLDNRNPDDLTSEYRLAETRGARSDHIPIPOHYWGWKV 346
Db 301 AEGAGYNGWTSGRQAQNLDNRNPDDLTSEYRLAETRGARSDHIPIPOHYWGWKV 357

RESULT 5
US-10-060-425-12
; Sequence 12, Application US/10060425
; Publication No. US20020164650A1
; GENERAL INFORMATION:
; APPLICANT: Hiebsch, Ronald
; TITLE OF INVENTION: Methods of Assessing Wolframin Protein Activity
; FILE REFERENCE: 00450.US1
; CURRENT APPLICATION NUMBER: US/10/060,425
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: 60/266,385
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 12
```

```
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-060-425-12

Query Match      30.4%; Score 1058; DB 13; Length 476;
Best Local Similarity 48.4%; Pred. No. 1.8e-89;
Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLADEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASRCARHARTYGI 202
Db 20 GWLLAAEQEPGAPAGMRRRLRLOQEDGISFEYHRYPELREALVSVWLQCTAISRIYTV 79
QY 203 GRSFDGRELIVIEFSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGN 262
Db 80 GRSFEGRELLVIELSDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEYQKGN 139
QY 263 PRIQRLNTRTHLLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDLNRNPPDLTSEYI 322
Db 140 ETIVNLHSTRIHIMPSLNPDGFKAASQPGELKDWFGVGRSNAQGDIDLNRNPPDLRIYV 199
QY 323 RLAEATRGARSDHI-----PIPOHYWGWKVAPETKAIKMKWMTIIPFVLSASLHGGDLVVS 376
Db 200 -VNEKEGGPNHLLKNLKKIVDQ---SKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255
QY 377 YPFDFSKHPOEKMFSPTPDEKMFKLLSRAYADVHPMMDRSENRCGGN-----FLKRGSG 431
Db 256 YPDETRSGTAHE-YSSCPDDAIFQSLARAYSFNPMVSDPNRPPCKNDSDSSFV--DG 312
QY 432 IINGADWYFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWOHNKESILNPFVETV 491
Db 313 TTNGGAWYSPVGGMQDFNYLSSNCFEITVELSCERFPPETUKSYWEDNKNLSINYLEQI 372
QY 492 HRGIGKGVTDKFKGPKVKNARISVKGIRHDITTPADGDYWRLLPPGHIHIVIAQAPGYAKVI 551
Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKODGYWRLLVPGNYKLTASAPGYLAIT 432
QY 552 KKVIIIPARMKRAGRVDFILQ 571
Db 433 KKVAVP--FSPAVGVDFELE 450

RESULT 7
US-09-827-040-2
; Sequence 2, Application US/09827040
; Patent No. US20010024792A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Meyer, Joanne
; TITLE OF INVENTION: Method of Detecting Risk of Type II Diabetes Based on
; FILE OF INVENTION: Mutations Found in Carboxypeptidase E
; FILE REFERENCE: 5800-14, 035800/174130
; CURRENT APPLICATION NUMBER: US/09/827,040
; CURRENT FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 09/233,989
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: 60/105,102
; PRIOR FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-827-040-2

Query Match      30.3%; Score 1054; DB 9; Length 476;
Best Local Similarity 48.2%; Pred. No. 4.3e-89;
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLADEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASRCARHARTYGI 202
Db 20 GWLLAAEQEPGAPAGMRRRLRLOQEDGISFEYHRYPELREALVSVWLQCTAISRIYTV 79
QY 203 GRSFDGRELIVIEFSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGN 262
Db 80 GRSFEGRELLVIELSDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEYQKGN 139
QY 263 PRIQRLNTRTHLLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDLNRNPPDLTSEYI 322
Db 140 ETIVNLHSTRIHIMPSLNPDGFKAASQPGELKDWFGVGRSNAQGDIDLNRNPPDLRIYV 199
QY 323 RLAEATRGARSDHI-----PIPOHYWGWKVAPETKAIKMKWMTIIPFVLSASLHGGDLVVS 376

Query Match      30.4%; Score 1057; DB 9; Length 476;
Best Local Similarity 48.4%; Pred. No. 2.3e-89;
Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLADEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASRCARHARTYGI 202
Db 20 GWLLAAEQEPGAPAGMRRRLRLOQEDGISFEYHRYPELREALVSVWLQCTAISRIYTV 79
QY 203 GRSFDGRELIVIEFSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGN 262
Db 80 GRSFEGRELLVIELSDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEYQKGN 139
QY 263 PRIQRLNTRTHLLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDLNRNPPDLTSEYI 322
Db 140 ETIVNLHSTRIHIMPSLNPDGFKAASQPGELKDWFGVGRSNAQGDIDLNRNPPDLRIYV 199
QY 323 RLAEATRGARSDHI-----PIPOHYWGWKVAPETKAIKMKWMTIIPFVLSASLHGGDLVVS 376
```

```
Db 200 -VNEKEGPNHLLKNMKKIVDQN---TKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255
Qy 377 YPFDPSKHQPEKMFSTPDEKMKLLSRAYADVHPMMDRSENRCGN-----FLKRGs 431
Db 256 YPYDETRSGSAHE-YSSSPDAIFQSLARAYSSFNPAUSDNPPCRKNDSSFFV--DG 312
Qy 432 IINGADWYSTGMSDNFYLHNCFEITVELGCVKPPPEALYTLQHNKESLNFVETV 491
Db 313 TTNGGAWSYVPGGMQDNFYLSSNCFEITVELSCFKPPPEETLKYTWEDNKNLSLYLEQI 372
Qy 492 HRGIGKVVYTKFGKPVGNARISVKGIHDDITTPADGDYWRLLPGIHIVIAQAPGAKVI 551
Db 373 HRGVKGVRDLQGNPIANATISVEGIDHDVTSKAGDGYWRLLIPGNYKLTAAPGYLAIT 432
Qy 552 KKVIIIPARMKRAGRVDFILO 571
Db 433 KKVAVP--YSPAAGVDPELE 450

RESULT 8
US-10-060-425-4
; Sequence 4, Application US/10060425
; Publication No. US20020164650A1
; GENERAL INFORMATION:
; APPLICANT: Hiebsch, Ronald
; TITLE OF INVENTION: Methods of Assessing Wolfram Protein Activity
; FILE REFERENCE: 00450.US1
; CURRENT APPLICATION NUMBER: US/10/060,425
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: 60/266,385
; PRIOR FILING DATE: 2001-02-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-425-4

Query Match 30.3%; Score 1054; DB 13; Length 476;
Best Local Similarity 48.2%; Pred. No. 4.3e-89;
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

Qy 155 GGLAEADALPSGLPPTFIR-----FSHHSYAQMVRVLRRTASCAHVARTYSI 202
Db 20 GLLLGAEAQEPGAPAGRRRRRLQEDGIGSFYHRYPELREALVSVWLQCTAISRIYTV 79
Qy 203 GRSPDGRELLVIEFSSRPGQHELMPEVKLIGNHNEVAGREMLIYLAQYLCSEYLLGN 262
Db 80 GRSPFEGRELLVIELSDNPGVHPEPEPFKYIGNHMGNEAVGRELLIFLAQYLCNEYQKGN 139
Qy 263 PRIQRLNTRTHLLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDLNRNFPDLTSEYY 322
Db 140 ETIVNLHSTRIHIMPSLNDPGFEKAASQPGELKDFVGRSNAQGGIDLNRNFPDLRIYV 199
Qy 323 RLAEATRGASDHI-----PIPOHYWGWKVPATKAIKMQMTIPFVLSASLHGGDLVVS 376
Db 200 -VNEKEGPNHLLKNMKKIVDQN---TKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255
Qy 377 YPFDPSKHQPEKMFSTPDEKMKLLSRAYADVHPMMDRSENRCGN-----FLKRGs 431
Db 256 YPYDETRSGSAHE-YSSSPDAIFQSLARAYSSFNPAUSDNPPCRKNDSSFFV--DG 312
Qy 432 IINGADWYSTGMSDNFYLHNCFEITVELGCVKPPPEALYTLQHNKESLNFVETV 491
Db 313 TTNGGAWSYVPGGMQDNFYLSSNCFEITVELSCFKPPPEETLKYTWEDNKNLSLYLEQI 372
Qy 492 HRGIGKVVYTKFGKPVGNARISVKGIHDDITTPADGDYWRLLPGIHIVIAQAPGAKVI 551
Db 373 HRGVKGVRDLQGNPIANATISVEGIDHDVTSKAGDGYWRLLIPGNYKLTAAPGYLAIT 432
Qy 552 KKVIIIPARMKRAGRVDFILO 571
```

```
Db 433 KKVAVP--YSPAAGVDPELE 450

RESULT 9
US-10-295-027-576
; Sequence 576, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 576
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-576

Query Match 30.3%; Score 1054; DB 14; Length 476;
Best Local Similarity 48.2%; Pred. No. 4.3e-89;
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

Qy 155 GGLAEADALPSGLPPTFIR-----FSHHSYAQMVRVLRRTASCAHVARTYSI 202
Db 20 GLLLGAEAQEPGAPAGRRRRRLQEDGIGSFYHRYPELREALVSVWLQCTAISRIYTV 79
Qy 203 GRSPDGRELLVIEFSSRPGQHELMPEVKLIGNHNEVAGREMLIYLAQYLCSEYLLGN 262
Db 80 GRSPFEGRELLVIELSDNPGVHPEPEPFKYIGNHMGNEAVGRELLIFLAQYLCNEYQKGN 139
Qy 263 PRIQRLNTRTHLLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDLNRNFPDLTSEYY 322
Db 140 ETIVNLHSTRIHIMPSLNDPGFEKAASQPGELKDFVGRSNAQGGIDLNRNFPDLRIYV 199
Qy 323 RLAEATRGASDHI-----PIPOHYWGWKVPATKAIKMQMTIPFVLSASLHGGDLVVS 376
Db 200 -VNEKEGPNHLLKNMKKIVDQN---TKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255
```

Qy	377	YPDFSKHQEKMFSPTDEKXFKLLSRAYDVHPMMORSNRCGN-----FLKRG8 431
Dd	256	YPIETRSGSAHE--YSSSPDDAIFOSLARAYSSEFNAMPDPNRPCPKNDSDSFV--DG 312
Qy	432	IINGADWYSFTGMSDENVLHNCNCFEITVELGCVKFPPEEALYTLWOHNKESLIINFVEY 491
Dd	313	TTWGGAWYSVPGMQDFNYLSNCPFEITVELSCBKFPEETLKTYWBEDNKNSIYSLEOI 372
Qy	492	HRIKGIVTDKFGPKVNARI8VKIGRIDHTTAPDGDYWRLLPPGIHVIAQAQPVAKVI 551
Dd	373	HRCVKGFVDQLGNFANATISVEGIDHDVTSAKDGDYWRLLIPENYKLTPASPGVIAT 432
Qy	552	KKVIIIPARMKRAGRVDFILO 571
Dd	433	KKVAVE--YSPAAGVDFFELE 450

RESULT 10  
US-10-723-860-720  
Sequence 720, Application US/10723860  
Publication No. US20040253606A1  
GENERAL INFORMATION:  
APPLICANT: Aziz, Natasha  
APPLICANT: Ginsburg, Wendy M.  
APPLICANT: Zlotnik, Albert  
TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
FILE REFERENCE: 05882, 0193.NPUS01  
CURRENT APPLICATION NUMBER: US/10/723,860  
CURRENT FILING DATE: 2003-11-26  
PRIOR APPLICATION NUMBER: 60/429,739  
PRIOR FILING DATE: 2002-11-26  
NUMBER OF SEQ ID NOS: 8393  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 720  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-723-860-720

Query Match	30.3%	Score 1054;	DB 17;
Best Local Similarity	48.2%	Pred. NO. 4.3e-89;	Length 476;
Matches 212;	Conservative	67;	Mismatches 129;
			Indels 32;
			Gaps 8;

[illegible]

; PRIOR APPLICATION NUMBER: 60/266,385  
; PRIOR FILING DATE: 2001-02-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Alysia  
US-10-060-425-15

Query Match 30.2%; Score 1052; DB 13; Length 434;  
Best Local Similarity 50.1%; Pred. No. 5.8e-89;  
Matches 206; Conservative 67; Mismatches 118; Indels 20; Gaps 7;  
Qy 172 IRFSHSHYAQMVRVLRRTASRCARHARTYIGRSFSGRELLVIEFFSRPGQHELMPEVK 231  
Db 7 ISPEYHRYPELREALVSVWLQCAVSRITYVGRSFEGRLLVIELSDNPGVHEPGEPEFK 66  
Qy 232 LIGNIHGNEAVGEMLIYLAQYLCEYLLGNPRIORLNTTRIHLPSINPDGVEVAAAE 291  
Db 67 YIGNMHNEAVGRELIFLAQYLCNEYQKGNETIVQLIHNTRIHIMPSLNDPGPEKAASQ 126  
Qy 292 GAGVNGWTSGRONAQLDLNRNPDLTSEYVRLAETRGARSDHI-----PIQHYWGWK 345  
Db 127 LGELKQWFGVRSNAQGLDLNRNPDLDRIYV-INEKEGGNNHLLKMKIVDQN---TK 182  
Qy 346 VAPETKAIKMMQTIPEVLSASLHGGDLVVSYPDFSKHPQEEKMFSPPTDEKMFKLLSR 405  
Db 183 LAPETKAVIHWINDIPEVLSANLHGGDLVANYPYDETRSGSAHE-YSSCPDDDAIFQSLAR 241  
Qy 406 AYADVHPMMDRSENRCGN-----FLKRSIINGADWYFTGMSDFNYLHNCPEITV 460  
Db 242 AYSSFNPMSPDRPPCRKNDSDSFEV--GTTNGAAWYSPVGMQDFNYLSSNCFEITV 299  
Qy 461 ELGCVKEPPEEALYTLQHNKESLLNFVETVHRGKGVVTDKFKPKVNARISVKGIRHD 520  
Db 300 ELSCKEPPEETLKNYWNEDNKNLSIYQIHRGVGVFVRDLQGNPIANATLSVEGIDHD 359  
Qy 521 ITTAPDGDYWRLLPPGHIHIVIAQAPGYAKVKKVIIIPARMKRAGRVDFILO 571  
Db 360 VTSKAGDYWRLLVPGNYKLTAAPGYLATKKVAVP--YSPAVRVDFELE 408

RESULT 13  
US-09-925-300-1715  
; Sequence 1715, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: PatentIn ver. 2.0  
; SEQ ID NO 1715  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (42)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (43)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (44)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1715  
Query Match 30.2%; Score 1051; DB 9; Length 491;  
Best Local Similarity 50.1%; Pred. No. 8.6e-89;  
Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;  
Qy 172 IRFSHSHYAQMVRVLRRTASRCARHARTYIGRSFSGRELLVIEFFSRPGQHELMPEVK 231  
Db 64 ISPEYHRYPELREALVSVWLQCAVSRITYVGRSFEGRLLVIELSDNPGVHEPGEPEFK 123  
Qy 232 LIGNIHGNEAVGEMLIYLAQYLCEYLLGNPRIORLNTTRIHLPSINPDGVEVAAAE 291  
Db 124 YIGNMHNEAVGRELIFLAQYLCNEYQKGNETIVNLHSTRIHIMPSLNDPGPEKAASQ 183  
Qy 292 GAGVNGWTSGRONAQLDLNRNPDLTSEYVRLAETRGARSDHI-----PIQHYWGWK 345  
Db 184 PGELKQWFGVRSNAQGLDLNRNPDLDRIYV-VNEKEGGNNHLLKMKIVDQN---TK 239  
Qy 346 VAPETKAIKMMQTIPEVLSASLHGGDLVVSYPDFSKHPQEEKMFSPPTDEKMFKLLSR 405  
Db 240 LAPETKAVIHWINDIPEVLSANLHGGDLVANYPYDETRSGSAHE-YSSCPDDDAIFQSLAR 298  
Qy 406 AYADVHPMMDRSENRCGN-----FLKRSIINGADWYFTGMSDFNYLHNCPEITV 460  
Db 299 AYSSFNPMSPDRPPCRKNDSDSFEV--GTTNGAAWYSPVGMQDFNYLSSNCFEITV 356  
Qy 461 ELGCVKEPPEEALYTLQHNKESLLNFVETVHRGKGVVTDKFKPKVNARISVKGIRHD 520  
Db 357 ELSCKEPPEETLKNYWNEDNKNLSIYQIHRGVGVFVRDLQGNPIANATLSVEGIDHD 416  
Qy 521 ITTAPDGDYWRLLPPGHIHIVIAQAPGYAKVKKVIIIPARMKRAGRVDFILO 571  
Db 417 VTSKAGDYWRLLVPGNYKLTAAPGYLATKKVAVP--YSPAGVDFELE 465

RESULT 14  
US-09-833-245-1035  
; Sequence 1035, Application US/09833245  
; Publication No. US20040010134A1  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Albumin Fusion Proteins  
; FILE REFERENCE: PF546PCT  
; CURRENT APPLICATION NUMBER: US/09/833,245  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: 60/229,358  
; PRIOR FILING DATE: 2000-04-12  
; PRIOR APPLICATION NUMBER: 60/256,931  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/199,384  
; PRIOR FILING DATE: 2000-04-25  
; NUMBER OF SEQ ID NOS: 2267  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1035  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (42)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (43)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (44)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-833-245-1035

Query Match 30.2%; Score 1051; DB 11; Length 491;  
Best Local Similarity 50.1%; Pred. No. 8.6e-89;  
Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;



```
QY 172 IRSHSHYAQMVRVLRRTASRCARHARTYSIGRSPDRELLVIERSSRPGQHLMPEVK 231
Db 64 ISFEYHRYDELREALVSWLQCTAISRIYTVGRSFEGRLLVIELSDNPGVHEPGEPEFK 123
QY 232 LIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIQRLNLTTRIHLPSINPDGYEVAAB 291
Db 124 YGNMGNEAVGRELLIFLAQYLCNEYQKNETIVNLIHSTRHIMPSPNPDGFEKAASQ 183
QY 292 GAGYNGWTSGRONALDNRNFPDLTSEYRLAETRGARSDHI-----PIQHYWGWK 345
Db 184 PGEKDFWFGSRNAQGDIDLRNFPDLDRIVY-VNEKEGGPNHLLKNMKKIVDQN---TK 239
QY 346 VAPETKAIMKMWOTIPFVLSASLHGGDLAVSVPPDFSKHPQEEKMFSPDEKMKLLSR 405
Db 240 LAPETKAVIHWIMDIPFVLSANLHGGDLVANYPDETRSGSAHE-YSSSPDDAIFQSLAR 298
QY 406 AYADVHPMMMDRSENRCGN-----FLKRGSIINGADWYSFTGGMDSDFNYLHTNCFEITV 460
Db 299 AYSSFNPAWSDPNRPPCRKNDSSFFV--DGTNGGAWYSVPGGQDFNYLSSNCFEITV 356
QY 461 ELGCVKFPPEALYTLWOHNKESLLNFVETVHRGKGVVTDKPGKPVKVARISVKGIRHD 520
Db 357 ELSCEKFPPEETLKTWEDNKNLSISYLEQIHRGVKGFVRDLQGNPIANATISVEGIDHD 416
QY 521 ITTAPDGDYWRLLPPGHIHIVIAQAPGYAKVIKKVIIIPARMKRAGRVDFILQ 571
Db 417 VTSKDGWYRLLIPGNYKLITASAPGYLAITKKVAVP--YSPAAGVDFFELE 465
```

Search completed: January 25, 2005, 09:05:02  
Job time : 152 secs

```
RESULT 15
US-09-833-245-1057
; Sequence 1057, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1057
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (42)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (44)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-1057
```

```
Query Match 30.2%; Score 1051; DB 11; Length 491;
Best Local Similarity 50.1%; Pred. No. 8.6e-89;
Matches 206; Conservative 66; Mismatches 119; Indels 20; Gaps 7;

QY 172 IRSHSHYAQMVRVLRRTASRCARHARTYSIGRSPDRELLVIERSSRPGQHLMPEVK 231
Db 64 ISFEYHRYDELREALVSWLQCTAISRIYTVGRSFEGRLLVIELSDNPGVHEPGEPEFK 123
QY 232 LIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIQRLNLTTRIHLPSINPDGYEVAAB 291
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2005, 08:35:14 ; Search time 200 Seconds  
(without alignments)  
1844.076 Million cell updates/sec

Title: US-10-757-262-104  
Perfect score: 3480  
Sequence: 1 MPPPPPLLLLTVLVVAARP.....WWSYFTSLTRPRLKY 641

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3480	100.0	641	2	O00520
2	2892.5	83.1	652	2	O54858
3	2887.5	83.0	652	2	O54859
4	2840.5	81.6	654	2	O8R4V4
5	2318	66.6	647	2	O8QGP3
6	1849.5	53.1	364	2	O96WX2
7	1057	30.4	476	1	CBPE_RAT
8	1054	30.3	476	1	CBPE_HUMAN
9	1052	30.2	434	1	CBPE_BOVIN
10	1052	30.2	454	2	O6NSM5
11	1052	30.2	454	2	AAH70026
12	1049	30.1	476	1	CBPE_MOUSE
13	1049	30.1	476	2	BAC33184
14	1041.5	29.9	450	2	O6NT94
15	1041.5	29.9	450	2	AAH66689
16	1036	29.8	734	2	O6P4G8
17	1036	29.8	734	2	O6UW65
18	1036	29.8	734	2	AAQ99315
19	1036	29.8	734	2	AAH63430
20	1035	29.7	454	1	CPXM_LOPAM
21	1028	29.5	734	1	CPXM_HUMAN
22	1010	29.0	433	2	O6PAV9
23	1010	29.0	453	2	AAH59995
24	1003	28.8	722	1	CPXM_MOUSE
25	1001.5	28.8	430	2	O8N2F1
26	1001	28.8	458	1	CBPN_HUMAN
27	997	28.6	457	2	O9J7N5
28	986.5	28.3	458	2	O91WN9
29	984	28.3	457	2	O9EQV8
30	961	27.6	756	1	CPX2_HUMAN
31	944	27.1	764	1	CPX2_MOUSE

RESULT 1

O00520	PRELIMINARY;	PRT;	641	AA.
AC	O00520;			
DT	01-JUL-1997 (TREMBlrel. 04, Created)			
DT	01-JUL-1997 (TREMBlrel. 04, Last sequence update)			
DT	01-MAR-2004 (TREMBlrel. 26, Last annotation update)			
DE	Carboxypeptidase Z precursor.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=97256770; PubMed=9099699;			
RA	Song L., Fricker L.D.;			
RT	"Cloning and expression of human carboxypeptidase Z, a novel			
RT	metallocarboxypeptidase."			
RL	J. Biol. Chem. 272:10543-10550(1997).			
DR	EMBL; U83411; AAB58911.1; -			
DR	HSSP; Q30240; 1H8L.			
DR	MEROPS; M14.012; -			
DR	Genew; HGNC:2333; CPZ.			
DR	GO; GO:0004181; P:metallocarboxypeptidase activity; TAS.			
DR	GO; GO:0006508; P:proteolysis and peptidolysis; TAS.			
DR	InterPro; IPR008969; Carboxypep_reg.			
DR	InterPro; IPR000024; Fz_domain.			
DR	InterPro; IPR008834; Peptidase_M14.			
DR	InterPro; IPR008575; Peptidase_M14B.			
DR	Pfam; PF05885; DUF857; 1.			
DR	Pfam; PF01392; Fz; 1.			
DR	Pfam; PF00246; Zn_carboxept; 1.			
DR	PRINTS; PR00765; CRBOXYPTASEA.			
DR	SMART; SM00063; FRI; 1.			
DR	SMART; SM00631; Zn_pept; 1.			
DR	PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.			
DR	PROSITE; PS00338; FZ; 1.			
KW	Carboxypeptidase; Signal.			
FT	SIGNAL 1 38 Potential.			
FT	CHAIN 19 641 carboxypeptidase Z.			
SQ	SEQUENCE 641 AA; 72503 MW; B229B5734E0AA94 CRC64;			

Query Match 100.0%; Score 3480; DB 2; Length 641;  
Best Local Similarity 100.0%; Pred. No. 1.2e-261;  
Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MPPPPPLLLLTVLVVAARPGCEFERNPAATCDVLQRLTCSDAAYNHTTFFPNLLQHSWE	60
Db	1	MPPPPPLLLLTVLVVAARPGCEFERNPAATCDVLQRLTCSDAAYNHTTFFPNLLQHSWE	60
QY	61	VVEASSEYILLSVLHOLLEGQCNPDRLILGCAVLAPRCGGWRRRCRHCIGLEVCQP	120
Db	61	VVEASSEYILLSVLHOLLEGQCNPDRLILGCAVLAPRCGGWRRRCRHCIGLEVCQP	120

32	906.5	26.0	1128	2	O88442	mus musculus
33	905	26.0	733	2	O6ZSC7	Q6ZSC7 homo sapien
34	905	26.0	733	2	BAC87026	BAC87026 homo sapi
35	904.5	26.0	719	2	O61281	mus musculus
36	895.5	25.7	845	2	O7KZ79	homo sapien
37	895.5	25.7	1158	2	Q14113	homo sapien
38	895.5	25.7	1158	2	O81UX7	Q81UX7 homo sapien
39	892	25.6	728	2	O97567	O97567 bos taurus
40	861	24.7	1389	2	O90240	O90240 anas sp. (d
41	858.5	24.7	1387	2	O57512	O57512 anas platyr
42	857.5	24.6	380	2	P83852	P83852 lophoneta
43	849.5	24.4	1380	2	O86X66	Q86X66 homo sapien
44	849	24.4	1380	2	O86SH9	Q86SH9 homo sapien
45	845	24.3	1380	1	CBPD_HUMAN	O75976 homo sapien

ALIGNMENTS

```

QY 121 AFDAIDMAWYFLDCHRYFTREDEGCYDPLEKRLGGLEADEALPSGLPTFFIRFSSHSA 180
DB 121 AFDAIDMAWYFLDCHRYFTREDEGCYDPLEKRLGGLEADEALPSGLPTFFIRFSSHSA 180
QY 181 QMVRVLRRTASRCARHARTYISGRSPDRELLVIEFSSRPGQHELMEPEVKLTGNHNE 240
DB 181 QMVRVLRRTASRCARHARTYISGRSPDRELLVIEFSSRPGQHELMEPEVKLTGNHNE 240
QY 241 VAGREMLIYLAQYLCSEYLLGNPRIORLLNTTRIHLPLSNPDGYEVAAAGAGYNGWTS 300
DB 241 VAGREMLIYLAQYLCSEYLLGNPRIORLLNTTRIHLPLSNPDGYEVAAAGAGYNGWTS 300
QY 301 GRQNAQNLNRPDPDLTSEYRLAETRGARSDHIPIQHYWVGKVPATKAIKMWQTI 360
DB 301 GRQNAQNLNRPDPDLTSEYRLAETRGARSDHIPIQHYWVGKVPATKAIKMWQTI 360
QY 361 PFVLSASLHGGDLVSYPPDFSKHPQEKMFSTPDEKMKLLSRAYADVHPMMDRSEN 420
DB 361 PFVLSASLHGGDLVSYPPDFSKHPQEKMFSTPDEKMKLLSRAYADVHPMMDRSEN 420
QY 421 RCGNLFKRGSIINGADWYFTGMSDFNYLHNTCFEITVELGCVKFPPEEALYTLWOHN 480
DB 421 RCGNLFKRGSIINGADWYFTGMSDFNYLHNTCFEITVELGCVKFPPEEALYTLWOHN 480
QY 481 KESLLNFVETVHRGKIGVWTDKFKPVKNARI SVKGIRHDIITAPDGYWRLLPGGHIV 540
DB 481 KESLLNFVETVHRGKIGVWTDKFKPVKNARI SVKGIRHDIITAPDGYWRLLPGGHIV 540
QY 541 IAQAPGYAKVIKKVILPARKRAGRVDFILOPLGMGPKNFHGLRTGPHDPLGGASSLG 600
DB 541 IAQAPGYAKVIKKVILPARKRAGRVDFILOPLGMGPKNFHGLRTGPHDPLGGASSLG 600
QY 601 EATEPDLRARROPSADGSKPWWWSYFTSLSTRPRWLLKY 641
DB 601 EATEPDLRARROPSADGSKPWWWSYFTSLSTRPRWLLKY 641

RESULT 2
O54858 PRELIMINARY; PRT; 652 AA.
AC O54858
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Carboxypeptidase Z.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Xin X., Fricker L.D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017637; AAC04668.1; -.
DR HSSP; Q90240; 1H8L.
DR MEROPS; M14.012; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004182; F:carboxypeptidase A activity; IEA.
DR GO; GO:0004888; F:transmembrane receptor activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0007275; P:development; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008969; Carboxypeptidase_reg.
DR InterPro; IPR000024; Fz domain.
DR InterPro; IPR008575; Peptidase_M14.
DR Pfam; PF05885; DUF857; 1.
DR Pfam; PF01392; Fz; 1.
DR PRINTS; PF00246; Zn_carboxypeptidase.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00063; FRI; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.

```

```

DR PROSITE; PS50038; FZ; 1.
KW Carboxypeptidase.
SQ SEQUENCE 652 AA; 73081 MW; BE085394B91A978C CRC64;

Query Match      83.1%; Score 2892.5; DB 2; Length 652;
Best Local Similarity 81.6%; Pred. No. 5.7e-216;
Matches 536; Conservative 42; Mismatches 58; Indels 21; Gaps 5;

QY 1 MPPPPILLILLTVLVA---AARPGCEFERNP-----AATCVDLQLTCSDAAYN 46
DB 1 MPTPTLLAALAAALAAVAAYPSCSPGDPDGKQORLASTSATCVDLHLRTCADAAYN 60
QY 47 HTTFPNLLQHRWEVVEASSEYLLSVLHOLLEGQCNPDRLRLGCAVLAPRCGGWVRP 106
DB 47 HTTFPNLLQHRWEVVEASSEYLLSVLHOLLEGQCNPDRLRLGCAVLAPRCGGWVRP 106
QY 61 HTSFTPLEHRSWEAVEASPEYLLGVLFLEGGQCNPDRLRLGCVLAPRCGGHTQRP 120
DB 61 HTSFTPLEHRSWEAVEASPEYLLGVLFLEGGQCNPDRLRLGCVLAPRCGGHTQRP 120
QY 107 CRHICEGLREVCQAPFADDAIDMAWYFLDCHRYFTREDEGCYDPLEKRLGGLEADEALPSG 166
DB 121 CRVCEGLREACQAPFADDAIDMAWYFLDCTQYFAPEEGCYDPLEQRLGDELVEEALPSG 180
QY 167 LPTTIRFSSHSAQWVRVLRRTASRCARHARTYISGRSPDRELLVIEFSSRPGQHELM 226
DB 181 LPTTIRFSAHSAQWVRVLRRTASRCARHARTYISGRSFEGLDVLVIEFSSRPGQHELM 240
QY 227 EPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIORLLNTTRIHLPLSNPDGYE 286
DB 241 EPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIORLLNTTRIHLPLSNPDGYE 300
QY 287 VAAAEAGAGYNGWTSGRQNAQNLNRPDPDLTSEYRLAETRGARSDHIPIQHYWVGKV 346
DB 301 VAAAEAGAGYNGWTSGRQNAQNLNRPDPDLTSEYRLAETRGARSDHIPIQHYWVGKV 360
QY 347 APETKAIKMWQTI PFVLSASLHGGDLVSYPPDFSKHPQEKMFSTPDEKMKLLSR 406
DB 361 APETKAIKMWQTI PFVLSASLHGGDLVSYPPDFSKHPQEKMFSTPDEKMKLLSR 420
QY 407 YADVHPMMDRSENRCGNFLKRGSIINGADWYFTGMSDFNYLHNTCFEITVELGCVK 466
DB 421 YADVHPMMDRSENRCGNFLKRGSIINGADWYFTGMSDFNYLHNTCFEITVELGCVK 480
QY 467 FPPPEALYTLWOHNKESLLNFVETVHRGKIGVWTDKFKPVKNARI SVKGIRHDIITAPD 526
DB 481 FPPPEALYTLWOHNKESLLNFVETVHRGKIGVWTDKFKPVKNARI SVKGIRHDIITAPD 540
QY 527 GDYWRLLPGLIHVIAQAPGYAKVIKKVILPARKRAGRVDFILOPLGMGPKNFHGLR 586
DB 541 GDYWRLLPGLIHVIAQAPGYAKVIKKVILPARKRAGRVDFILOPLGMGPKNFHGLR 600
QY 587 TGPH--DPLGGASSLGATEPDLRARROPSADGSKPWWWSYFTSLSTRPRWLLKY 641
DB 601 ALPRSLDPQGAQAL----DFEPPPRARQP-ASGSKPWWWSYFTSLSTRPRWLLKY 652

RESULT 3
O54859 PRELIMINARY; PRT; 652 AA.
AC O54859
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Carboxypeptidase Z.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Xin X., Fricker L.D.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017638; AAC04669.1; -.
DR HSSP; Q90240; 1H8L.
DR MEROPS; M14.012; -.

```

DR GO: GO:0016020; C:membrane; IEA.  
DR GO: GO:0004182; F:carboxypeptidase A activity; IEA.  
DR GO: GO:0004888; F:transmembrane receptor activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0007275; P:development; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR008969; Carboxypeptidase\_reg.  
DR InterPro: IPR000024; Fz\_domain.  
DR InterPro: IPR00834; Peptidase\_M14.  
DR InterPro: IPR008575; Peptidase\_M14B.  
DR Pfam: PF05885; DUF857; 1.  
DR Pfam: PF01392; Fz; 1.  
DR Pfam: PF00246; Zn\_carboxypeptidase.  
DR PRINTS: SM00631; Zn\_pept; 1.  
DR SMART: SM00631; Zn\_pept; 1.  
DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; 1.  
DR PROSITE: PS50038; Fz; 1.  
KW Carboxypeptidase.  
SQ SEQUENCE 652 AA; 73071 MW; F3AB539434B9F19B CRC64;  
  
Query Match 83.0%; Score 2887.5; DB 2; Length 652;  
Best Local Similarity 81.9%; Pred. No. 1.4e-215;  
Matches 538; Conservative 41; Mismatches 57; Indels 21; Gaps 7;  
  
QY 1 MPPPPLLL-----LLTVLVVAA---ARPG-----CE-FERNPAATCVDLQRLTCSDAAYN 46  
DB 1 MPTPTLLLAALAAALAAVAAYSSCGPDPSPGKQCRQLASTHATCVDLHLRTCADAAYN 60  
  
QY 47 HTTFPPLLQHRSEVVEASSEYILLVHLLGQCNPDLRLGLCAVLAPRCGGWVR 106  
DB 61 HTSFPTPLEHRSWEAVESEYLLGVLFLLGQCNPDLRLGLGCVLAPRCGGHTQ 120  
  
QY 107 CRHCEGLREVCOPAFDAIDMAMPYFLDCHRYFTREDEGCDYDPLEKLRGLGDEALPSG 166  
DB 121 CRVCEGLREACOPAFDAIDMAMPYFLDCTQYFAPEEGCYDPLEQURGDELVEALPSG 180  
  
QY 167 LPPTFIRFHSYQAQVVRVLRRTASRCAHVARTYSIGRSFGDRELAVIEFSRPGQHELM 226  
DB 181 LPPTFIRFHSYQAQVVRVLRRTAARCSQVAKTYSIGRSFEGKDLVIEFSRPGQHELM 240  
  
QY 227 EPEVKLIGNHNEVAGREMLIYLAQVLCSEYLLGNPRIQLNTRTHLLPSINPDGYE 286  
DB 241 EPEVKLIGNHNEVAGREMLIYLAQVLCSEYLLGNPRIQLNTRTHLLPSINPDGYE 300  
  
QY 287 VAAEAGAGYNGWTSGRQNAQNLNRRNFPDLTSEYVRLAETRCARGSDHIPPIQHYWGVK 346  
DB 301 VAAEAGAGYNGWTSGRQNAQNLNRRNFPDLTSEYVRLAETRCARGSDHIPPIQHYWGVK 360  
  
QY 347 APETKAIKMKWQITIPFVLSASLHGGDLVVSYPDFSKHPQEEKMFSPDPDEKMFKLLSRA 406  
DB 361 APETKAIKMKWQITIPFVLSASLHGGDLVVSYPDFSKHPQEEKMFSPDPDEKMFKLLARA 420  
  
QY 407 YADVHPMDRSENRCGNLFRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGCVK 466  
DB 421 YADVHPMDRSENRCGNLFRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGCVK 480  
  
QY 467 FPPEEALYTLWQNKESLNFVTHRGIKGVVTDKFGKPKVKNARI SVKGIHDIITAPD 526  
DB 481 FPPEEALYTLWQNKESLNFVTHRGIKGVVTDKFGKPKVKNARI SVKGIHDIITAPD 540  
  
QY 527 GDYWRLLPPGIHIVIAQAPGYAKVKKVITIPARMKGRAGRVDFILQPLGMPKPNFIHGLRR 586  
DB 541 GDYWRLLPPGIHIVIAQAPGYAKVKKVITIPARMKGRAGRVDFILQPLGMPKPNFIHGLRR 600  
  
QY 587 TGGH--DPLGGASLSGATEPDLRARRQPSADGSKPWWWSYFTSLSTRHPRLLKY 641  
DB 601 ALPRSLDPQAPAL-----DFEPPRRARQP-AGSKPWWWSYFTSLSPHPRLLKY 652  
  
PRELIMINARY; PRT; 654 AA.

RESULT 4

Q8R4V4

ID Q8R4V4

AC Q8R4V4;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Carboxypeptidase Z.  
GN Name=Cpz;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10050;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NMRI;  
RL Moeller C., Swindell E.C., Eichele G.;  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF356844; AA00219.1; -  
DR HSSP: Q90240; 1H8L.  
DR MEROPS: M14.012; -.  
DR MGD: MGI:88487; Cpz.  
DR GO: GO:0016020; C:membrane; IEA.  
DR GO: GO:0004182; F:carboxypeptidase A activity; IEA.  
DR GO: GO:0004888; F:transmembrane receptor activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0007275; P:development; IEA.  
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro: IPR008969; Carboxypeptidase\_reg.  
DR InterPro: IPR000024; Fz\_domain.  
DR InterPro: IPR00834; Peptidase\_M14.  
DR InterPro: IPR008575; Peptidase\_M14B.  
DR Pfam: PF05885; DUF857; 1.  
DR Pfam: PF01392; Fz; 1.  
DR Pfam: PF00246; Zn\_carboxypeptidase.  
DR PRINTS: PR00765; CRBOXYPTASEA.  
DR SMART: SM00631; Zn\_pept; 1.  
DR SMART: SM00631; Zn\_pept; 1.  
DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; 1.  
DR PROSITE: PS50038; Fz; 1.  
KW Carboxypeptidase.  
SQ SEQUENCE 654 AA; 73694 MW; E3C52E0CA8A7D65D CRC64;  
  
Query Match 81.6%; Score 2840.5; DB 2; Length 654;  
Best Local Similarity 80.2%; Pred. No. 6.3e-212;  
Matches 530; Conservative 47; Mismatches 57; Indels 27; Gaps 7;  
  
QY 1 MPPPPLLL-----LLTVLVVAA---ARPGCEFERNP-----AATCVDLQRLTCSDA 44  
DB 1 MPTPTLLLAALAAALAAVAALARPSPSCSPGDPSPGKQRLVYTHSATCVDLHLRTCADAA 60  
  
QY 45 YNHTTTPNLLQHRSEVVEASSEYILLVHLLGQCNPDLRLGLCAVLAPRCGGWVR 104  
DB 61 YNHTSPTPLEHRSWEAVESEYLLGVLFLLGQCNPDLRLGLGCVLAPRCGGHTQ 120  
  
QY 105 RPCRHI CEGREVCOPAFDAIDMAMPYFLDCHRYFTREDEGCDYDPLEKLRGLGDEALP 164  
DB 121 RPCRHI CEGREVCOPAFDAIDMAMPYFLDCAQYFAPEEGCYDPLEELRGELDVEALA 180  
  
QY 165 SGLPPTFIRFHSYQAQVVRVLRRTASRCAHVARTYSIGRSFGDRELAVIEFSRPGQHE 224  
DB 181 SGLPPTFIRFHSYQAQVVRVLRRTAARCSQVAKTYSIGRSFEGKDLVIEFSRPGQHE 240  
  
QY 225 LMEPEVKLIGNHNEVAGREMLIYLAQVLCSEYLLGNPRIQLNTRTHLLPSINPDG 284  
DB 241 LMEPEVKLIGNHNEVAGREMLIYLAQVLCSEYLLGNPRIQLNTRTHLLPSINPDG 300  
  
QY 285 YEVAAGAGYNGWTSGRQNAQNLNRRNFPDLTSEYVRLAETRCARGSDHIPPIQHYWVG 344  
DB 301 YEVAAGAGYNGWTSGRQNAQNLNRRNFPDLTSEYVRLAETRCARGSDHIPPIQHYWVG 360  
  
QY 345 KVAPETKAIKMKWQITIPFVLSASLHGGDLVVSYPDFSKHPQEEKMFSPDPDEKMFKLLS 404  
DB 361 KVAPETKAIKMKWQITIPFVLSASLHGGDLVVSYPDFSKHPQEEKMFSPDPDEKMFKLLA 420  
  
QY 405 RAYADVHPMDRSENRCGNLFRGSIINGADWYSFTGMSDFNYLHNTNCFEITVELGC 464



RA Fujimori Y., Komiya M., Tashiro H., Tanigami A., Fujiwara T.,  
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
 RA Kawabata A., Hikiuji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
 RA Togauchi T., Oyama M., Hata H., Watanabe M., Komatsu T.,  
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
 RT "Complete sequencing and characterization of 21,243 full-length human  
 RT cDNAs.";  
 RL Nat. Genet. 36:40-45 (2004).  
 DR EMBL; AK056317; BAB71147.1; -.  
 DR HSP; Q90240; IHL8.  
 DR MEROPS; M14.012; -.  
 DR CO; GO:0016020; C-membrane; IEA.  
 DR CO; GO:0004182; F:carboxypeptidase A activity; IEA.  
 DR CO; GO:0004180; F:carboxypeptidase activity; IEA.  
 DR CO; GO:0004888; F:transmembrane receptor activity; IEA.  
 DR CO; GO:0007275; F:development; IEA.  
 DR CO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR00024; Pz domain.  
 DR InterPro; IPR000834; Peptidase\_M14.  
 DR Pfam; PF01392; Fz; 1.  
 DR Pfam; PF00246; Zn\_carbopept; 1.  
 DR PRINTS; PR00765; CRBOXYPTASEA.  
 DR SMART; SM00631; FRI; 1.  
 DR SMART; SM00631; Zn\_pept; 1.  
 DR PROSITE; PS50038; Fz; 1.  
 DR Carboxypeptidase.  
 KW Carboxypeptidase.  
 SQ SEQUENCE 364 AA; 41216 MW; A3AF492A75362299 CRC64;  
 Query Match 53.1%; Score 1849.5; DB 2; Length 364;  
 Best Local Similarity 96.4%; Pred. No. 3.2e-135;  
 Matches 344; Conservative 1; Mismatches 1; Indels 11; Gaps 1;  
 QY 1 MPPPPPLLLLTVLVAAARCGCFERNPA-----ATCVDLQLRTCSDAAYNHTT 49  
 DB 1 MPPPPPLLLLTVLVAAARCGCFERNPAGECHRPAAADSATCVDLQLRTCSDAAYNHTT 60  
 QY 50 FPNLLQHRSEWVEASSEYILLVHLLGQCNPDRLAGCAVLAPRCGGWVRPCR 109  
 DB 61 FPNLLQHRSEWVEASSEYILLVHLLGQCNPDRLAGCAVLAPRCGGWVRPCR 120  
 QY 110 ICEGLREVCOPAFDAIDMAMPYFLDCHRYETREDEGCDPLEKRLGLEADEALPSGLPP 169  
 DB 121 ICEGLREVCOPAFDAIDMAMPYFLDCHRYETREDEGCDPLEKRLGLEADEALPSGLPP 180  
 QY 170 TFIHFSSHVAQWVRLRRTASCAHVARTYSIGRSFDRGLLVIEFSSRPGQHELMPE 229  
 DB 181 TFIHFSSHVAQWVRLRRTASCAHVARTYSIGRSFDRGLLVIEFSSRPGQHELMPE 240  
 QY 230 VKLIGNTHGNEVAGREMLIYLAQVLCSEYLLGNPRIQLNTRILLPSINPDGYEVA 289  
 DB 241 VKLIGNTHGNEVAGREMLIYLAQVLCSEYLLGNPRIQLNTRILLPSINPDGYEVA 300  
 QY 290 AEGAGYNGWTSGRNAQNLNRPDLTSEYTLAETRGARSDHPIPIQHYWNGKV 346  
 DB 301 AEGAGYNGWTSGRNAQNLNRPDLTSEYTLAETRGARSDHPIPIQHYWNGKV 357

## RESULT 7

CBPE RAT

ID CBPE RAT STANDARD; PRT; 476 AA.

AC P15087;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Carboxypeptidase E precursor (EC 3.4.17.10) (CPH) (Carboxypeptidase H)

DE (CPE) (Enkephalin convertase) (Prohormone processing

DE carboxypeptidase).

GN Name=Cpe;

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=101116;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RA MEDLINE=89174664; PubMed=2784437;  
 RA Rodriguez C., Brayton K.A., Brownstein M., Dixon J.E.;  
 RT "Rat procarboxypeptidase H. Cloning, characterization, and sequence  
 RT of the cDNA and regulation of the mRNA by corticotropin-releasing  
 RT factor.";  
 RL J. Biol. Chem. 264:5988-5995 (1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89261823; PubMed=2755530;  
 RA Fricker L.D., Adelman J.P., Douglass J., Thompson R.C.;  
 RA von Strandmann R.P., Hutton J.;  
 RT "Isolation and sequence analysis of cDNA for rat carboxypeptidase E  
 RT [EC 3.4.17.10], a neuropetide processing enzyme.";  
 RL Mol. Endocrinol. 3:666-673 (1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Brain;  
 RX MEDLINE=90241164; PubMed=2334405;  
 RA Manser E., Fernandez D., Loo L., Goh P.Y., Monfries C., Hall C.,  
 RA Lim L.;  
 RT "Human carboxypeptidase E. Isolation and characterization of the cDNA,  
 RT sequence conservation, expression and processing in vitro.";  
 RL Biochem. J. 267:517-525 (1990).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92123221; PubMed=1770952;  
 RA Jung Y.K., Kunczt C.J., Pearson R.K., Dixon J.E., Fricker L.D.;  
 RT "Structural characterization of the rat carboxypeptidase-E gene.";  
 RL Mol. Endocrinol. 5:1257-1268 (1991).  
 CC -I- FUNCTION: Removes residual C-terminal Arg or Lys remaining after  
 CC initial endoprotease cleavage during prohormone processing.  
 CC -I- CATALYTIC ACTIVITY: Release of C-terminal arginine or lysine  
 CC residues from polypeptides.  
 CC -I- SUBCELLULAR LOCATION: Secretory granules of pancreatic islets,  
 CC adrenal gland, pituitary and brain.  
 CC -I- SIMILARITY: Belongs to peptidase family M14.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; J04625; AAA40875.1; -.  
 CC EMBL; M31602; AAA40873.1; -.  
 CC EMBL; X51406; CAA35768.1; -.  
 CC EMBL; L07281; AAA40957.1; ALT SEQ.  
 CC EMBL; L07273; AAA40957.1; JOINED.  
 CC EMBL; L07274; AAA40957.1; JOINED.  
 CC EMBL; L07275; AAA40957.1; JOINED.  
 CC EMBL; L07277; AAA40957.1; JOINED.  
 CC EMBL; L07278; AAA40957.1; JOINED.  
 CC EMBL; L07279; AAA40957.1; JOINED.  
 CC EMBL; L07280; AAA40957.1; JOINED.  
 CC PIR; A40469; A40469.  
 CC PIR; S12461; S12461.  
 CC HSP; Q90240; IHL8.  
 CC MEROPS; M14.005; -.  
 CC RGD; 2394; Cpe.  
 CC InterPro; IPR008969; Carboxypep reg.  
 CC InterPro; IPR008934; Peptidase\_M14.  
 CC InterPro; IPR008575; Peptidase\_M14B.  
 CC Pfam; PF05885; DUF857; 1.  
 CC Pfam; PF00246; Zn\_carbopept; 1.  
 CC PRINTS; PR00765; CRBOXYPTASEA.  
 CC SMART; SM00631; Zn\_pept; 1.

```

DR PROSITE; PS00132; CARBOXYPEPT ZN 1; 1.
DR DR PROSITE; PS00133; CARBOXYPEPT ZN 2; 1.
KW Carboxypeptidase; Glycoprotein; Hydrolase; Metalloprotease; Signal;
KW Zinc; Zymogen.
FT SIGNAL 1 27 Or 34.
FT PROPEP 28 42 Activation peptide.
FT CHAIN 43 476 Carboxypeptidase E.
FT METAL 114 114 Zinc (By similarity).
FT METAL 117 117 Zinc (By similarity).
FT METAL 248 248 Zinc (By similarity).
FT ACT_SITE 342 342 Nucleophile (By similarity).
FT CARBOHYD 139 139 N-linked (GlcNAc. .) (Potential).
FT CARBOHYD 390 390 N-linked (GlcNAc. .) (Potential).
FT CONFLICT 82 82 S -> T (in Ref. 2).
FT CONFLICT 415 415 V -> A (in Ref. 3).
FT CONFLICT 453 453 S -> Y (in Ref. 3).
SQ SEQUENCE 476 AA; 53309 MW; C2213D1DFDECA120 CRC64;

Query Match 30.4%; Score 1057; DB 1; Length 476;
Best Local Similarity 48.4%; Pred. No. 1.9e-73;
Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLAEADALPSGLPTPIR-----FSHHSYAQMVRVLRRTARCAHVARTYGI 202
Db 20 GWLLAAEAQEPGAPAAAGMRRRLRQEDGISTFEYHYPRLREALVSVWLQCTAISRIYTV 79
QY 203 GRSFGRELLVIEFSSRPGOHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEVLLGN 262
Db 80 GRSFEGRELLVIELSDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCSEYQRCN 139
QY 263 PRIQLLNTTRIHLPSINPDGVEAAARGAGYNGWTSGRQNAQNLDRNPNPDLTSEYY 322
Db 140 ETVNLHSTRHIMESLPDGFKAASQPGELKDFWGRSNAQGLDNRNPFLLDRIVY 199
QY 323 RLAETRGASDHI-----PIPOHYWKGKVPETKAIKMKWQTIPTVLSASLHGGLVVS 376
Db 200 -VNEKEGGPNHLLKNLKVQDN---SKLAPETKAVIHWIMDIPVLSANLHGGDLVAN 255
QY 377 YPFDPSKHQEEKWFSPTDEKFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431
Db 256 YPDETRSGTAHE-YSSCPDDAIFQSLARAYSFNPNVSDPNRPCKRKNDDSSFY--DG 312
QY 432 IINGADWYFTGMSDFNLHTNCFEITVELGCVKPPPEALYTLWQHNKESILLNFVETV 491
Db 313 TTNGGAWYSPGCMQDFNYLSSNCFEITVELSCEKPEETLKSIEDKNLSLNLVLEQI 372
QY 492 HRGIKGVTDKFKPKVNARISVKGIRHDITAPDGDYWRLLPPGHIVIAQPGYAKVI 551
Db 373 HRGKGVFVRDLQGNPIANATISVDGIDHDVTSKADGDYWRLLVPGNYKLTASAPGYLAIT 432
QY 552 KKVLIIPARMKRAGRVDFILQ 571
Db 433 KKVAVP--FSPAGVDFELE 450

RESULT 8
CBPE_HUMAN STANDARD; PRT; 476 AA.
AC Fl6870; Q9UIU9;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE Carboxypeptidase E precursor (EC 3.4.17.10) (CPH) (Carboxypeptidase H)
DE (CPE) (Enkephalin convertase) (Prohormone processing
DE carboxypeptidase).
DE Name=CPE;
DE OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;

```

```

RX MEDLINE=90241164; PubMed=2334405;
RA Manser E., Fernandez D., Loo L., Goh P.Y., Monfries C., Hall C.,
RA Lim L.;
RT "Human carboxypeptidase E. Isolation and characterization of the cDNA,
RT sequence conservation, expression and processing in vitro.";
RA Biochem. J. 267:517-525(1990).
RX SEQUENCE FROM N.A.
RX MEDLINE=98324430; PubMed=9662053;
RA Utsunomiya N., Ohagi S., Sanke T., Tatsuta H., Hanabusa T., Nanjo K.;
RT "Organization of the human carboxypeptidase E gene and molecular
RT scanning for mutations in Japanese subjects with NIDDM or obesity.";
RA Diabetologia 41:701-705(1998).
RX SEQUENCE FROM N.A.
RX TISSUE=Colon, and Eye;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.R., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Hellon E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Removes residual C-terminal Arg or Lys remaining after
CC initial endoprotease cleavage during prohormone processing.
CC Processes proinsulin.
CC -!- CATALYTIC ACTIVITY: Release of C-terminal arginine or lysine
CC residues from polypeptides.
CC -!- SUBCELLULAR LOCATION: Secretory granules of pancreatic islets,
CC adrenal gland, pituitary and brain.
CC -!- SIMILARITY: Belongs to peptidase family M14.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X51405; CAA35767.1; -
CC EMBL; AB006898; BAA86053.1; -
CC EMBL; AB006899; BAA86053.1; JOINED.
CC EMBL; AB006890; BAA86053.1; JOINED.
CC EMBL; AB006892; BAA86053.1; JOINED.
CC EMBL; AB006893; BAA86053.1; JOINED.
CC EMBL; AB006894; BAA86053.1; JOINED.
CC EMBL; AB006895; BAA86053.1; JOINED.
CC EMBL; AB006896; BAA86053.1; JOINED.
CC EMBL; AB006897; BAA86053.1; JOINED.
CC EMBL; BC033866; AAH33866.1; -
CC EMBL; BC033612; AAH33612.1; -
CC PIR; S09489; S09489.
CC HSSP; Q90240; 1H81.
CC MEROPS; M14.005; -.
CC Genew; HGNC:2303; CPE.
CC MIM; 114855; -.
CC GO; GO:0005886; C:plasma membrane; TAS.
CC GO; GO:0004180; F:carboxypeptidase activity; TAS.

```



DR GO: 0008152; P: metabolism; TAS.  
 DR InterPro: IPR008969; Carboxypeptid reg.  
 DR InterPro: IPR00834; Peptidase\_M14.  
 DR InterPro: IPR008575; Peptidase\_M14B.  
 DR Pfam: PF05885; DUF857; 1.  
 DR Pfam: PF00246; Zn carbopet; 1.  
 DR PRINTS: PR00765; CRBOXYPTASEA.  
 DR PROSITE: PS00132; CARBOXYPEPT\_ZN\_1; 1.  
 DR PROSITE: PS00133; CARBOXYPEPT\_ZN\_2; 1.  
 KW Carboxypeptidase; Glycoprotein; Hydrolase; Metalloprotease; Signal;  
 ZINC; Zymogen.  
 FT SIGNAL 1 25 Potential.  
 FT PROPEP 26 42 Activation peptide.  
 FT CHAIN 43 476 Carboxypeptidase E.  
 FT METAL 114 114 Zinc (By similarity).  
 FT METAL 117 117 Zinc (By similarity).  
 FT METAL 248 248 Zinc (By similarity).  
 FT METAL 342 342 Nucleophile (By similarity).  
 FT ACT SITE 342 342 Nucleophile (By similarity).  
 FT CARBOHYD 139 139 N-linked (GlcNAc...) (Potential).  
 FT CARBOHYD 390 390 N-linked (GlcNAc...) (Potential).  
 FT CONFLICT 391 391 A -> R (in Ref. 2).  
 SQ SEQUENCE 476 AA; D561AC0285A51E86 CRC64;  
  
 Query Match 30.3%; Score 1054; DB 1; Length 476;  
 Best Local Similarity 48.2%; Pred. No. 3.2e-73;  
 Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;  
  
 QY 155 GGLAEALPSGLPPTPIR-----FSHSHYAQMVRVLRRTASRCAHVARTYSI 202  
 DB 20 GWLGLAEAPGAPAGARRRRRLQEDGGSFYHRYHREALVSVWLQCTAISRIYV 79  
  
 QY 203 GRSGDRELIVBFSSRPGOHELMPEVEVKLIHGNHGNVAGREMLIYLAQYLCSEYLLGN 262  
 DB 80 GRSGFEGRELLIVELSDNPGVHPECEPEFKYIGNHMGNEAVGRELLIFLAQYLCNEYQKGN 139  
  
 QY 263 PRIQLRLNTHRIHLLPSINPDGVEVAEAGVNGWTSGRQNAQLDLNRPDLTSEYY 322  
 DB 140 ETIVNLTHSTRIHPSLNPDGFEKAASQPGELKDWFGVGRSNAQGLDLNRPDLDRIVY 199  
  
 QY 323 RLAETRGASDHI-----PIPHYWMGVKAPETKAIKMKMTIPFVLSASLHGGDLVVS 376  
 DB 200 -VNEKEGPNHLLKMKKIVDQN---TKLAPETKAVIHWIMDIPFVLSANLHGGDLVAN 255  
  
 QY 377 YPFDFSKHPOEEMFSPDPEKFKLLSRAYADVHPMMMDRSNRCGN-----FLKRG 431  
 DB 256 YPYDETRSGSAHE-YSSSPDPAIFQSLARAYSSFNPA MSPNPPCKRKNDDSSFV--DG 312  
  
 QY 432 IINGADWYSTGMSDNFNYLHNCFEITVBLGCVKPPPEALYTLQHNKESLLNFVETV 491  
 DB 313 TTNGGAWYSPGGMQDFNYLSSNCFEITVELSCFKFPPEETLKYTWEDNKNLSIYLEQI 372  
  
 QY 492 HRGIGKGVVTKFGKPKVNARISVKGIRHDTTAPDGDYWELELPGIHIVTAQAPYAKVI 551  
 DB 373 HRGKGVFRDLQGNPNANATISVEGIDHDVTSKQGDYWELELPGNYKLITASAPGYLAIT 432  
  
 QY 552 KKVIIIPARKKRAGRVDFILQ 571  
 DB 433 KKVAVP--YSPAGVDFELE 450  
  
 RESULT 9  
 CBPE\_BOVIN STANDARD; PRT; 434 AA.  
 AC P04836;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Carboxypeptidase B (EC 3.4.17.10) (CPH) (Carboxypeptidase H) (CPE)  
 DE (Enkephalin convertase) (Prohormone processing carboxypeptidase).  
 GN Name=CPE;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.  
 NCBI\_TaxID=9913;  
 [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=87014809; PubMed=3020433;  
 Fricker L.D., Evans C.J., Esch F.S., Herbert E.;  
 "Cloning and sequence analysis of cDNA for bovine carboxypeptidase  
 E.";  
 Nature 323:461-464 (1986).  
 -!- FUNCTION: Removes residual C-terminal Arg or Lys remaining after  
 initial endoprotease cleavage during prohormone processing.  
 -!- CATALYTIC ACTIVITY: Release of C-terminal arginine or lysine  
 residues from polypeptides.  
 -!- SUBCELLULAR LOCATION: Secretory granules of pancreatic islets,  
 adrenal gland, pituitary and brain.  
 -!- SIMILARITY: Belongs to peptidase family M14.  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@sib-sib.ch).  
 EMBL; X04411; CAA27999.1; -.  
 PIR; A24327; A24327.  
 HSSP; Q90240; 1H8L.  
 MEROPS; M14.005; -.  
 InterPro: IPR008969; Carboxypeptid reg.  
 InterPro: IPR00834; Peptidase\_M14.  
 InterPro: IPR008575; Peptidase\_M14B.  
 Pfam: PF05885; DUF857; 1.  
 Pfam: PF00246; Zn carbopet; 1.  
 PRINTS; PR00765; CRBOXYPTASEA.  
 SMART; SM00631; Zn pept; 1.  
 PROSITE; PS00132; CARBOXYPEPT\_ZN\_1; 1.  
 PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; 1.  
 Carboxypeptidase; Glycoprotein; Hydrolase; Metalloprotease; Zinc.  
 METAL 72 72 Zinc (By similarity).  
 METAL 75 75 Zinc (By similarity).  
 ACT SITE 206 206 Zinc (By similarity).  
 CARBOHYD 97 97 Nucleophile (By similarity).  
 CARBOHYD 348 348 N-linked (GlcNAc...) (Potential).  
 SEQUENCE 434 AA; 49205 MW; E1E848FE5139DBED CRC64;  
  
 Query Match 30.2%; Score 1052; DB 1; Length 434;  
 Best Local Similarity 50.1%; Pred. No. 4.1e-73;  
 Matches 206; Conservative 67; Mismatches 118; Indels 20; Gaps 7;  
  
 QY 172 IRFSHSHYAQMVRVLRRTASRCAHVARTYSIGRSFDGRELIVLEFSSRPGOHELMPEVK 231  
 DB 7 ISFESHRYHREALVSVWLQCAAVSRITVGRSGFEGRELLVLELSDNPGVHPEGPEFK 66  
  
 QY 232 LTGINHGNHGNVAGREMLIYLAQYLCSEYLLGNPRIQLRLNTHRIHLLPSINPDGVEVAEAE 291  
 DB 67 YIGNHGNHGNVAGREMLIYLAQYLCSEYLLGNPRIQLRLNTHRIHLLPSINPDGVEVAEAE 126  
  
 QY 292 GAGYNGWTSGRQNAQLDLNRPDLTSEYYRLAETRGASDHI-----PIPHYWMGVK 345  
 DB 127 LGELKDWFGVGRSNAQGLDLNRPDLDRIVY-INEKEGPNHLLKMKKIVDQN---TK 182  
  
 QY 346 VAPETKAIKMKMTIPFVLSASLHGGDLVSVFPDFSKHPOEEMFSPDEKMKFLLSR 405  
 DB 183 LAPETKAVIHWIMDIPFVLSANLHGGDLVANPYDETRSGSAHE-YSSCDDDDIFOSLAR 241  
  
 QY 406 AYADVHPMMMDRSNRCGN-----FLKRGSIINGADWYSTGMSDNFNYLHNCFEITV 460  
 DB 242 AYSSFNPPMSDPRPCKRKNDDSSFVE--GITNGAWYSPGGMQDFNYLSSNCFEITV 299  
  
 QY 461 ELGCVKFPPEALYTLQHNKESLLNFVETVHRGIGKGVVTKFGKPKVNARISVKGIRH 520

```

Db 300 ELSCEKPEETLKNYWNEDKNKSLISYIQIHRGVKGFVRDLQGNPIANATLSVEGIDHD 359
QY 521 ITTAPDGDYWRLLPPGHIHIVIAQPGYAKVIKKVITPARMKRAGRVDPILO 571
Db 360 VTSAGODGYWRLLVPGNYKLTASAPGLAIKAVAP--YSPAVRVDPELE 408

RESULT 10
Q6NSM5
ID Q6NSM5 PRELIMINARY; PRT; 454 AA.
AC Q6NSM5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ZGC:85981.
GN Name=zgc:85981;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Peters G.J., Abramson R.D., Mullany S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gunaratne P.H.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Whitings R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Blakesley R.W., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Smallos D.E., Schnerch A., Schein J.E.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RP "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC070026; AAH70026.1; -.
DR InterPro; IPR008969; Carboxypeptid reg.
DR InterPro; IPR00834; Peptidase_M14.
DR InterPro; IPR008575; Peptidase_M14B.
DR Pfam; PF05885; DUF857; 1.
DR PRINTS; PR00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 454 AA; 51282 MW; 269DD657BCB24C79 CRC64;

Query Match 30.2%; Score 1052; DB 2; Length 454;
Best Local Similarity 50.0%; Pred. No. 4.4e-73;
Matches 206; Conservative 68; Mismatches 120; Indels 18; Gaps 7;

QY 172 IRFSSHSYAQMVRVLRRTASRCAHVARTSYIGRSFDGRELIVTFSSRPGQHLMPEVK 231
Db 27 ISFEVHYRVEEMRKSLVSVMLQCPSTIRITVGVSEFGRRELLVLENSDNGIHEPGEPEFK 86
QY 232 LIGNTHGNVAGREMLIYLAQVLCSEYLLGNPRIORLLNTTHILLPINDPGYVAAB 291
Db 87 YIGNHGNVAGRELLIYLAQVLCNEYOQEGNDTIIDLTHSTRIHIMPNSMNPDFEKAASQ 146

```

```

QY 292 GAGYNGWTSGRQNAQLDNRFNPDLTSEYVLAETRGARSDHI-----PIQHYWVGK 345
Db 147 PGEMKDWFGVSRNAQOIGIDLRNFPDLDRIVY--MNEREGANNHLLKNMKKAVDEN--TK 202
QY 346 VAPETKAIMKWMQITIPVLSASLHGGDLVSVPPDFSKHPQEKMFSTPDEKMPKLSR 405
Db 203 LAPETKAVIHWIMDIPFVLSANLHGGDVVANYPYDETRSGSTHE--YSASPDPLVFKSLAK 261
QY 406 AYADVHPMMDRSENRCGN-----FLKRGSIINGADWYSFTGCMGDFNYLHNCPEITVE 461
Db 262 AYSIYNPVMSDTRPPCKNDDSSFKEG--ITNGAWYSVPGMQDFNYSNCEFTILE 320
QY 462 LGCVPPEPEALYTLWQNKESLLNFVETVHRGKGVVTDKFGKVPQANRISVKGIRHDI 521
Db 321 LSCDKFPPEDSLKQYWDQNRNSLVNIIEQVHRGVSGFVRDLQGNIPISNASVSEGIDHDI 380
QY 522 TTAPDGDYWRLLPPGHIHIVIAQPGYAKVIKKVITPARMKRAGRVDPILOPL 573
Db 381 TTKAGDYWRLLAPNGYKVSASAPGYLTIVVKKVAVP--HSPSTRLDLFELES 430

```

## RESULT 11

```

AAH70026
ID AAH70026 PRELIMINARY; PRT; 454 AA.
AC AAH70026;
DT 01-JUN-2004 (TrEMBLrel. 27, Created)
DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Brownstein M.J., Ussin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Peters G.J., Abramson R.D., Mullany S.J.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gunaratne P.H.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Whitings R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Blakesley R.W., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Smallos D.E., Schnerch A., Schein J.E.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RP "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC070026; AAH70026.1; -.
DR InterPro; IPR008969; Carboxypeptid reg.
DR InterPro; IPR00834; Peptidase_M14.
DR InterPro; IPR008575; Peptidase_M14B.
DR Pfam; PF05885; DUF857; 1.
DR PRINTS; PR00246; Zn_carboxypept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ SEQUENCE 454 AA; 51282 MW; 269DD657BCB24C79 CRC64;

Query Match 30.2%; Score 1052; DB 2; Length 454;
Best Local Similarity 50.0%; Pred. No. 4.4e-73;
Matches 206; Conservative 68; Mismatches 120; Indels 18; Gaps 7;

QY 172 IRFSSHSYAQMVRVLRRTASRCAHVARTSYIGRSFDGRELIVTFSSRPGQHLMPEVK 231

```

Db 27 ISFYHRYEENKSLVSWLQCPSTIRYTVGESFEGRELLVLEMSNPNHPEGEPEFK 86  
QY 232 LIGNIHGNEVAGREMLIYLAQYLCEYLLGNPRIQRLNTRIHLPSINPDGVEVAAS 291  
Db 87 YIGNMHGNEAVGRELLIYLAQYLCEYLLGNPRIQRLNTRIHLPSINPDGVEVAAS 146  
QY 292 GAGYNGWTSGRQNAQNLNRPDLTSEYVRLAETGARDHIT-----PIPOHYWGWK 345  
Db 147 PGMKDFVGRSNAQIGDLNRPDLTSEYVRLAETGARDHIT-----PIPOHYWGWK 202  
QY 346 VAPETKAIMKMQIPVLSLHGGDLVSVYPPDFSKHQEKKMFPSTDEKFKLLSR 405  
Db 203 LAPETKAVIHWMDIPVLSLHGGDLVSVYPPDFSKHQEKKMFPSTDEKFKLLSR 261  
QY 406 AYADVHPMMWDRSENRCGNN-----FLKRGSTINGADWYFTGGMDFNYLHTNCFEITVE 461  
Db 262 AYSYINPVMSDTNPPCKKDDSSFKEG-ITNGGANYSVPGQMDNYLSSNCFEITLE 320  
QY 462 LGCVKFPPEALYTLWQHNKESLNFVETVHRGKIGVYVDFKGPVKNARISVKGIRHDI 521  
Db 321 LSCDKFPPEALYTLWQHNKESLNFVETVHRGKIGVYVDFKGPVKNARISVKGIRHDI 380  
QY 522 TTAPDGDYWRLLPGLIHVIAQAPGYAKYKVKVILPARKRAGRVDPIQLPL 573  
Db 381 TTAKDGDYWRLLPGLIHVIAQAPGYAKYKVKVILPARKRAGRVDPIQLPL 430

## RESULT 12

CBPE\_MOUSE STANDARD; PRT; 476 AA.  
AC Q00493; Q64439;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE Carboxypeptidase B precursor (EC 3.4.17.10) (CPH) (Carboxypeptidase H)  
DE (CPE) (Enkephalin convertase) (Prohormone processing  
DE carboxypeptidase).  
GN Name=Cpe;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL Submitted (AUG-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A., AND VARIANT PRO-244.  
RC STRAIN=HRS/J;  
RX MEDLINE=95392564; PubMed=7663508;  
RA Naggett J.K., Pricker L.D., Varlamov O., Nishina P.M., Rouille Y.,  
RA Steiner D.F., Carroll R.J., Paigen B.J., Leiter E.H.;  
RT "Hyperproinsulinemia in obese fat/fat mice associated with a  
RT carboxypeptidase B mutation which reduces enzyme activity";  
RL Nat. Genet. 10:135-142(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Mammary gland;  
RX MEDLINE=22386257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grove L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: Removes residual C-terminal Arg or Lys remaining after  
CC initial endoprotease cleavage during prohormone processing.  
CC Processes proinsulin.  
CC -1- CATALYTIC ACTIVITY: Release of C-terminal arginine or lysine  
CC residues from polypeptides.  
CC -1- SUBCELLULAR LOCATION: Secretory granules of pancreatic islets,  
CC adrenal gland, pituitary and brain.  
CC -1- DISEASE: Defects in Cpe are the cause of the fat phenotype. Mice  
CC homozygous for the fat mutation develop obesity and hyperglycaemia  
CC that can be suppressed by treatment with exogenous insulin.  
CC -1- SIMILARITY: Belongs to peptidase family M14.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X61232; CAA43550.1; -;  
DR EMBL; U23184; AAB60488.1; -;  
DR EMBL; BC010197; AAI10197.1; -;  
DR PIR; S16383; S16383.  
DR HSSP; Q90240; 1H8L.  
DR MEROPS; M14.005; -;  
DR MGD; MGI:101932; Cpe.  
DR InterPro; IPR008969; Carboxypeptidase reg.  
DR InterPro; IPR008934; Peptidase\_M14.  
DR InterPro; IPR008575; Peptidase\_M14B.  
DR Pfam; PF00885; DUF857; 1.  
DR Pfam; PF00246; Zn\_carboxypeptidase; 1.  
DR PRINTS; PR00765; GRBOXYPTASEA.  
DR PROSITE; PS00132; CARBOXYPEPT\_ZN\_1; 1.  
DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; 1.  
DR Carboxypeptidase; Disease mutation; Glycoprotein; Hydrolase;  
KW Metalloprotease; Signal; Zinc; Zymogen.  
FT SIGNAL 1 27 By similarity.  
FT PROPEP 28 42 Activation peptide (By similarity).  
FT CHAIN 43 476 Carboxypeptidase E.  
FT METAL 114 114 Zinc (By similarity).  
FT METAL 117 117 Zinc (By similarity).  
FT METAL 248 248 Zinc (By similarity).  
FT ACT\_SITE 342 342 Nucleophile (By similarity).  
FT CARBOHYD 139 139 N-linked (GlcNAc...) (Potential).  
FT CARBOHYD 390 390 N-linked (GlcNAc...) (Potential).  
FT VARIANT 244 244 S -> P (in hyperproinsulinemia obese fat/  
FT fat mice; reduced activity).  
FT CONFLICT 25 25 A -> R (in Ref. 1).  
FT CONFLICT 81 81 R -> A (in Ref. 1).  
FT CONFLICT 167 167 S -> W (in Ref. 1).  
FT CONFLICT 171 171 E -> Q (in Ref. 1).  
FT CONFLICT 301 301 R -> C (in Ref. 1).  
FT CONFLICT 310 310 V -> D (in Ref. 1).  
FT CONFLICT 334 334 S -> SS (in Ref. 1).  
FT CONFLICT 344 344 S -> T (in Ref. 1).  
FT CONFLICT 367 367 S -> N (in Ref. 1).  
FT CONFLICT 388 388 I -> N (in Ref. 1).  
FT CONFLICT 425 425 A -> D (in Ref. 1).  
SQ SEQUENCE 476 AA; 53255 MW; 82CS318A12A17567 CRC64;  
Query Match 30.1%; Score 1049; DB 1; Length 476;  
Best Local Similarity 48.0%; Pred. No. 7.9e-73;  
Matches 211; Conservative 67; Mismatches 130; Indels 32; Gaps 8;  
QY 155 GGLAEALPSGLPPTFIR-----FHHSHYQWVRVLRRTASCAHVARTYSI 202

Db 20 GLLLTAEAEQGAAGRRRRRLQQEDGISFEVHYRPELREALVSWLQCTAISRIYTV 79  
 QY 203 GRSPDRELLVIEFSSRPGQHELMPEVVKLIGNHNEVAGREMLIYLAQVLCSEYLLGN 262  
 Db 80 GRSPGRELIVIELSDNPGVHEPGEPEFKYIGNHMGNEAVGRELLIFLAQVLCNEQKGN 139  
 QY 263 PRIQRLNLTTRIHLPSINPDGYVAAEAGYNGWTSGRQNAQNLDNENFDPDLTSEY 322  
 Db 140 ETIVNLHSTRIHMPSLNPDGFKAASQGEKDFWFGVRSNAQGDILNRFNPDRLRIVY 199  
 QY 323 RLAEETRGAARDHI-----PIPOHYWVGKVPETKAIMKWMQITPFVLSASLHGGDLVVS 376  
 Db 200 -VNEKEGPNHLLKNLKIYDQN---SKLAPEKAVIHWIMDIPFVLSANLHGGDLVAN 255  
 QY 377 YPFDFSKHPQBEKMFSTPEDEKPKLISRAYADVHPMMDRSENRCGN-----FLKRG 431  
 Db 256 YPYDETRSGTAHE-YSSCPDDAIFQSLARAYSSFNPMVSDPNRPPCKRKNDDSSFV--DG 312  
 QY 432 IINGADWYFTGMSDFNYLHNTNCFEITVELGCVKPPPEALATLWQHNKESLNFVETV 491  
 Db 313 TTNGGAWYSPVGGMDFNYLSSNCFEITVELSCEKFPPEETLSYWNEDNKNLSIYLEQI 372  
 QY 492 HRGIGKGVVTDKFGKPVKNARISVKGIRHDITTAAPDGYWRLLPPGHIHIVIAQAAGYAKVI 551  
 Db 373 HRGKGFVRDLQGNPIANATISVDGIDHVTSAKGDYWLLAPGNVYKLTASAPGYLAIT 432  
 QY 552 KKVLIIPARKKAGRVDFILO 571  
 Db 433 KKVAVP--FSPAVGVDFELE 450

## RESULT 13

BAC33184  
 ID BAC33184 PRELIMINARY; PRT; 476 AA.  
 AC BAC33184;  
 DT 14-APR-2004 (TrEMBLrel. 27, Created)  
 DT 14-APR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 14-APR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE 16 days embryo head cDNA, RIKEN full-length enriched library,  
 DE clone.C130020F14 product:carboxypeptidase E, full insert  
 DE sequence.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690 (2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.,  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44 (1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=20499374; PubMed=11042159;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630 (2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa S., Ohara E., Wakahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.,  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771 (2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Head;  
 RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,  
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,  
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozawa T.,  
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,  
 RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,  
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,  
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,  
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,  
 RA Tagawa A., Takahashi F., Takaku-Akai H., Tanaka Y., Tanaka T.,  
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK047891; BAC33184.1; -.  
 KW Carboxypeptidase.  
 SQ SEQUENCE 476 AA; 53255 MW; 82C5318A12A17567 CRC64;

## Query Match

30.1%; Score 1049; DB 2; Length 476;

Best Local Similarity 48.0%; Pred. No. 7.9e-73;

Matches 211; Conservative 67; Mismatches 130; Indels 32; Gaps 8;

QY 155 GGLEADEALPSGLPPTFIR-----FSHSVAQMVRLRRRTASCAHVARTYSI 202  
 Db 20 GLLLTAEAEQGAAGRRRRRLQQEDGISFEVHYRPELREALVSWLQCTAISRIYTV 79  
 QY 203 GRSPDRELLVIEFSSRPGQHELMPEVVKLIGNHNEVAGREMLIYLAQVLCSEYLLGN 262  
 Db 80 GRSPGRELIVIELSDNPGVHEPGEPEFKYIGNHMGNEAVGRELLIFLAQVLCNEQKGN 139  
 QY 263 PRIQRLNLTTRIHLPSINPDGYVAAEAGYNGWTSGRQNAQNLDNENFDPDLTSEY 322  
 Db 140 ETIVNLHSTRIHMPSLNPDGFKAASQGEKDFWFGVRSNAQGDILNRFNPDRLRIVY 199  
 QY 323 RLAEETRGAARDHI-----PIPOHYWVGKVPETKAIMKWMQITPFVLSASLHGGDLVVS 376  
 Db 200 -VNEKEGPNHLLKNLKIYDQN---SKLAPEKAVIHWIMDIPFVLSANLHGGDLVAN 255  
 QY 377 YPFDFSKHPQBEKMFSTPEDEKPKLISRAYADVHPMMDRSENRCGN-----FLKRG 431  
 Db 256 YPYDETRSGTAHE-YSSCPDDAIFQSLARAYSSFNPMVSDPNRPPCKRKNDDSSFV--DG 312  
 QY 432 IINGADWYFTGMSDFNYLHNTNCFEITVELGCVKPPPEALATLWQHNKESLNFVETV 491  
 Db 313 TTNGGAWYSPVGGMDFNYLSSNCFEITVELSCEKFPPEETLSYWNEDNKNLSIYLEQI 372  
 QY 492 HRGIGKGVVTDKFGKPVKNARISVKGIRHDITTAAPDGYWRLLPPGHIHIVIAQAAGYAKVI 551  
 Db 373 HRGKGFVRDLQGNPIANATISVDGIDHVTSAKGDYWLLAPGNVYKLTASAPGYLAIT 432  
 QY 552 KKVLIIPARKKAGRVDFILO 571  
 Db 433 KKVAVP--FSPAVGVDFELE 450

Query Match	29.9%;	Score 1041.5;	DB 2;	Length 450;
Best Local Similarity	50.5%;	Pred. No. 2.8e-72;		



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2005, 08:43:09 ; Search time 43 Seconds  
(without alignments)  
1434.301 Million cell updates/sec

Title: US-10-757-262-104  
Perfect score: 3480  
Sequence: 1 MPDPPLLLTTLVLAAR.....WWSYFTSLTHRPWLKY 641

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1058	30.4	476	2 S12461	carboxypeptidase E
2	1057	30.4	476	2 A40469	carboxypeptidase E
3	1054	30.3	476	2 S09489	carboxypeptidase E
4	1052	30.2	448	2 A24327	carboxypeptidase E
5	1035	29.7	454	2 A54324	carboxypeptidase H
6	1019	29.3	477	2 S16383	carboxypeptidase E
7	1001	28.8	458	2 S02074	lysine carboxypept
8	904.5	26.0	719	2 S51739	transcription repr
9	895.5	25.7	845	2 JC5256	adipocyte transcr
10	861	24.7	1389	2 I50090	carboxypeptidase g
11	823.5	23.7	1446	2 T30916	carboxypeptidase D
12	819	23.5	472	2 T20454	hypothetical prote
13	776.5	22.3	985	2 T29910	hypothetical prote
14	750	21.6	1404	2 T13420	probable carboxype
15	743.5	21.4	1119	2 T13284	carboxypeptidase (
16	743.5	21.4	1406	2 T13211	probable carboxype
17	707	20.3	439	2 A32619	carboxypeptidase M
18	705	20.3	501	2 T25343	hypothetical prote
19	630.5	18.1	491	2 B96739	hypothetical prote
20	253	7.3	424	2 S17571	carboxypeptidase T
21	234	6.7	451	2 S20723	carboxypeptidase (
22	193.5	5.6	999	2 T36021	probable zinc-bind
23	191	5.5	581	2 JC7086	F2D10 protein - hu
24	183.5	5.3	565	2 JE0338	Frizzled-2 protein
25	183	5.3	666	2 T24170	hypothetical prote
26	180.5	5.2	1113	2 JE0315	low-density lipopr
27	176	5.1	355	2 C83850	gamma-D-glutamyl-L
28	175	5.0	317	2 JE0175	frizzled protein-1
29	173.5	5.0	666	2 JC7312	frizzled-3 protein

30	168	4.8	574	2 JE0339	Frizzled-7 protein
31	162	4.7	647	2 JE0337	Frizzled-1 protein
32	161.5	4.6	295	2 JE0174	frizzled protein-2
33	161	4.6	579	2 JC7629	membrane-type friz
34	161	4.6	706	2 JE0164	frizzled-6 protein
35	159.5	4.6	588	2 T25162	Frizzled-1 protein
36	156	4.5	694	2 S71786	wingless receptor
37	151.5	4.4	537	2 JC7127	frizzled protein 4
38	151	4.3	351	2 G84128	hypothetical prote
39	147.5	4.2	373	2 T26030	hypothetical prote
40	146.5	4.2	558	2 S29125	dimethylaniline mo
41	144.5	4.2	641	2 A45054	probable intercell
42	144	4.1	197	2 JC7735	frizzled-related p
43	143	4.1	581	2 S03540	gene frizzled prot
44	142	4.1	417	2 A32128	carboxypeptidase A
45	137.5	4.0	550	2 T37325	wingless protein r

ALIGNMENTS

RESULT 1

S12461  
carboxypeptidase E - human  
C;Species: Homo sapiens (man)  
C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
C;Accession: S12461  
R;Hall, C.  
submitted to the EMBL Data Library, January 1990  
A;Reference number: S12461  
A;Accession: S12461  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-476 <HAL>  
A;Cross-references: UNIPROT:P15087; EMBL:X51406; NID:g55870; PID:g55871  
C;Superfamily: human carboxypeptidase H

Query Match 30.4%; Score 1058; DB 2; Length 476;  
Best Local Similarity 48.4%; Pred. No. 6.7e-76;  
Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

Qy	155	GGLEADALPSGLPTTFR-----FSHSHYAQMVRVLRRTASTCAHVARTYSI	202
Db	20	GWLLAAEAQEPGAPAAAGMRRRLRLQEDGISFEYHRYPELRREALVSVMLQCTAISRIYTV	79
Qy	203	GRSFDGRELIVIEFSRRPGOHELMREVKLIHNGHVEAGREMLIYLAOYLCEYLLGN	262
Db	80	GRSFDGRELIVIELSNPGVHPEPEFKYIGNMHGNEAVGRELLIFLAOYLCNEYQKGN	139
Qy	263	PRIORLLNTTRIHLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDNENFPDLTSEY	322
Db	140	ETIVNLIHSTRIHIMSLNPDGFEKAASQPGELKDMFVGRSNAQGLDNRNFPDLDRIVY	199
Qy	323	RLAETRGARSDHI-----PIPOHYMWKQVAPETKAIMKMQTIPFVLSASLHGGDLVVS	376
Db	200	-VNEKEGGPNHLLKNLKIIVDQN---SKLAPETKAVIHIMDIPFVLSANLHGGDLVAN	255
Qy	377	YPDPFSKHQEEKMFSPETDEKMPKLLSRAYADVHPHMDRSENRCGN-----FLKRG	431
Db	256	YPDETGRSGTAHE-YSSCPDDAIFQSLARAYSNFVMSDPNRPCKRKNDDSSFV--DG	312
Qy	432	IINGADWYSPFGMSDFNYLHTNCFEITVELGCVKPPPEALYTLWQHKNESLNFVETV	491
Db	313	TTNGGAWYSPGGNQDFNYLSSNCFEITVELSCEKPPPEETLKSIEDNKNLSNLYEQI	372
Qy	492	HRGIGKGVTDKFGKPVKNARISVKGRHDIITAPDGDYWRLLPPGHIVIAQAPGYAKVI	551
Db	373	HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGDWELLAPGNYKLTASAPGLAIT	432
Qy	552	KKVIVIPARMKGRGVDFILQ	571
Db	433	KKVAVP--FSPAVGVDFELE	450



RESULT 2  
A0469 carboxypeptidase B (EC 3.4.17.10) precursor - rat  
N/Alternate names: carboxypeptidase B  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 14-Feb-1992 #sequence revision 14-Feb-1992 #text\_change 09-Jul-2004  
C/Accession: A0469; A32871; A40154; S09490; A61258  
R/Jung, Y.K.; Kunczt, C.J.; Pearson, R.K.; Dixon, J.E.; Fricker, L.D.  
Mol. Endocrinol. 5, 1257-1268, 1991  
A/Title: Structural characterization of the rat carboxypeptidase-B gene.  
A/Reference number: A0469; MUID:92123221; PMID:1770952  
A/Accession: A0469  
A/Molecule type: DNA  
A/Residues: 1-476 <UN>  
A/Cross-references: UNIPROT:P15087; GB:L07273  
R/Rodriguez, C.; Brayton, K.A.; Brownstein, M.; Dixon, J.E.  
J. Biol. Chem. 264, 5988-5995, 1989  
A/Title: Rat procarboxypeptidase H. Cloning, characterization, and sequence of the cDNA  
A/Reference number: A32871; MUID:89174664; PMID:2784437  
A/Accession: A32871  
A/Molecule type: mRNA  
A/Residues: 1-476 <OD>  
A/Cross-references: GB:J04625; NID:G203303; PIDN:AAA40875.1; PID:G203304  
R/Fricker, L.D.; Adelman, J.P.; Douglass, J.; Thompson, R.C.; von Strandmann, R.P.; Hutt  
Mol. Endocrinol. 3, 666-673, 1989  
A/Title: Isolation and sequence analysis of cDNA for rat carboxypeptidase B (EC 3.4.17.1)  
A/Reference number: A40154; MUID:89261823; PMID:2725530  
A/Accession: A40154  
A/Molecule type: mRNA  
A/Residues: 1-81, 'T', 83-476 <PRI>  
A/Cross-references: GB:M31602; NID:G203296; PIDN:AAA40873.1; PID:G203297  
R/Manser, E.; Fernandez, D.; Loo, L.; Goh, P.Y.; Monfries, C.; Hall, C.; Lim, L.  
Biochem. J. 267, 517-525, 1990  
A/Title: Human carboxypeptidase E. Isolation and characterization of the cDNA, sequence  
A/Reference number: S09489; MUID:90241164; PMID:2334405  
A/Accession: S09489  
A/Molecule type: mRNA  
A/Residues: 1-7, 'G', 9-414, 'A', 416-452, 'Y', 454-476 <MAN>  
A/Cross-references: EMBL:X51406  
A/Note: the authors translated the codon CCG for residue 7 as Ser, GGG for residue 8 as  
R/Castano, L.; Russo, E.; Zhou, L.; Lipes, M.A.; Eisenbarth, G.S.  
J. Clin. Endocrinol. Metab. 73, 1197-1201, 1991  
A/Title: Identification and cloning of a granule autoantigen (carboxypeptidase-H) associ  
A/Reference number: A61258; MUID:92064702; PMID:1955501  
A/Accession: A61258  
A/Status: not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 200-335 <CAS>  
C/Superfamily: human carboxypeptidase H  
C/Keywords: hydrolase; metallo-carboxypeptidase; zymogen  
F/1-34/Domain: signal sequence #status predicted <SIG>  
F/35-42/Domain: activation peptide #status predicted <ACT>  
F/43-476/Product: carboxypeptidase H #status predicted <MAT>  
Query Match 30.4%; Score 1057; DB 2; Length 476;  
Best Local Similarity 48.4%; Pred. No. 8e-76;  
Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;  
QY 155 GGLAEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHVARTYSI 202  
DB 20 GLLAAEAQEPGAPAGMRRRLRQEDGISFEYHRYPELREALVSWLQCTAISRIYTV 79  
QY 203 GRSPDRELLVIEFSSRPGQHELMPEVKLIIGNHNEVAGREMLIYLAQYLCSEYLLGN 262  
DB 80 GRSPDRELLVIELSDNPGVHEPGEPEFYIGNHNEVAGREMLIYLAQYLCSEYQKGN 139  
QY 263 PRIQLINTRIHLLPSINPDGEVAAAGAGYNGWTSGRQNAQNLDLRNPPDLTSEY 322  
DB 140 ETIVNLHSTRIHMLSPDGEFKAASQPGELKDFVGRSNAQIGIDLNRNPPDLRIYV 199  
QY 323 RLAEATRGARSDHI-----PIPOHYWGWKVPETKAIKMKMOTIPFVLSASLHGGDLVVS 376  
DB 200 -VNEKEGGPNHLLKNMKKIVDQN---TKLAPETKAVIHWMIDIPFVLSANLHGGDLVAN 255  
QY 377 YPFDFSKHPQEKMFSPPTDEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
DB 256 YPYDETRSGSAHE--YSSSPDDAIFQSLARAYSSFNPMSPDNRPPCKNDDSSFV--DG 312  
QY 432 INGADWTSFTCGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHKNESLLNFVETV 491  
DB 80 GRSPDRELLVIELSDNPGVHEPGEPEFYIGNHNEVAGREMLIYLAQYLCSEYQKGN 139  
QY 263 PRIQLINTRIHLLPSINPDGEVAAAGAGYNGWTSGRQNAQNLDLRNPPDLTSEY 322  
DB 140 ETIVNLHSTRIHMLSPDGEFKAASQPGELKDFVGRSNAQIGIDLNRNPPDLRIYV 199  
QY 323 RLAEATRGARSDHI-----PIPOHYWGWKVPETKAIKMKMOTIPFVLSASLHGGDLVVS 376  
DB 200 -VNEKEGGPNHLLKNMKKIVDQN---TKLAPETKAVIHWMIDIPFVLSANLHGGDLVAN 255  
QY 377 YPFDFSKHPQEKMFSPPTDEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
DB 256 YPYDETRSGSAHE--YSSSPDDAIFQSLARAYSSFNPMSPDNRPPCKNDDSSFV--DG 312  
QY 432 INGADWTSFTCGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHKNESLLNFVETV 491  
DB 313 TTNGGAWTSVPGGMDFNYLSSNCFEITVELSCFEPPEETLKTYWEDNKNLSLYLQI 372  
QY 492 HRIKGVVTDKFGKPKVNARISVKGIHDIITAPDGDYWRLLPPGHIHIVIAQAPGAKVI 551  
DB 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGYWRLLVPGNYKLTASAPGYLAIT 432

Db 200 -VNEKEGGPNHLLKNKKIVDQN---SKLAPETKAVIHWMIDIPFVLSANLHGGDLVAN 255  
QY 377 YPFDFSKHPQEKMFSPPTDEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
Db 256 YPYDETRSGTAHE--YSSCPDDAIFQSLARAYSSFNPMSPDNRPPCKNDDSSFV--DG 312  
QY 432 IINGADWTSFTCGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHKNESLLNFVETV 491  
Db 313 TTNGGAWTSVPGGMDFNYLSSNCFEITVELSCFEPPEETLKTYWEDNKNLSLYLQI 372  
QY 492 HRIKGVVTDKFGKPKVNARISVKGIHDIITAPDGDYWRLLPPGHIHIVIAQAPGAKVI 551  
Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGYWRLLVPGNYKLTASAPGYLAIT 432  
QY 552 KKVIIIPARMKRAGRVDFILQ 571  
Db 433 KKVAVP--FSPAVGVDFELE 450  
RESULT 3  
S09489 carboxypeptidase B (EC 3.4.17.10) precursor - human  
N/Alternate names: carboxypeptidase E  
C/Species: Homo sapiens (man)  
C/Date: 12-Feb-1993 #sequence revision 12-Feb-1993 #text\_change 09-Jul-2004  
C/Accession: S09489  
R/Manser, E.; Fernandez, D.; Loo, L.; Goh, P.Y.; Monfries, C.; Hall, C.; Lim, L.  
Biochem. J. 267, 517-525, 1990  
A/Title: Human carboxypeptidase E. Isolation and characterization of the cDNA, sequence  
A/Reference number: S09489; MUID:90241164; PMID:2334405  
A/Accession: S09489  
A/Molecule type: mRNA  
A/Residues: 1-476 <MAN>  
A/Cross-references: UNIPROT:P16870; EMBL:X51405; NID:G29666; PIDN:CAA35767.1; PID:G29667  
C/Genetics:  
A/Gene: GDB:CPE  
A/Cross-references: GDB:127894; OMIM:114855  
A/Map position: 4pter-4qter  
C/Superfamily: human carboxypeptidase H  
C/Keywords: hydrolase; metallo-carboxypeptidase  
Query Match 30.3%; Score 1054; DB 2; Length 476;  
Best Local Similarity 48.2%; Pred. No. 1.4e-75;  
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;  
QY 155 GGLAEALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHVARTYSI 202  
Db 20 GLLAAEAQEPGAPAGMRRRLRQEDGISFEYHRYPELREALVSWLQCTAISRIYTV 79  
QY 203 GRSPDRELLVIEFSSRPGQHELMPEVKLIIGNHNEVAGREMLIYLAQYLCSEYLLGN 262  
Db 80 GRSPDRELLVIELSDNPGVHEPGEPEFYIGNHNEVAGREMLIYLAQYLCSEYQKGN 139  
QY 263 PRIQLINTRIHLLPSINPDGEVAAAGAGYNGWTSGRQNAQNLDLRNPPDLTSEY 322  
Db 140 ETIVNLHSTRIHMLSPDGEFKAASQPGELKDFVGRSNAQIGIDLNRNPPDLRIYV 199  
QY 323 RLAEATRGARSDHI-----PIPOHYWGWKVPETKAIKMKMOTIPFVLSASLHGGDLVVS 376  
Db 200 -VNEKEGGPNHLLKNMKKIVDQN---TKLAPETKAVIHWMIDIPFVLSANLHGGDLVAN 255  
QY 377 YPFDFSKHPQEKMFSPPTDEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
Db 256 YPYDETRSGSAHE--YSSSPDDAIFQSLARAYSSFNPMSPDNRPPCKNDDSSFV--DG 312  
QY 432 INGADWTSFTCGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHKNESLLNFVETV 491  
Db 313 TTNGGAWTSVPGGMDFNYLSSNCFEITVELSCFEPPEETLKTYWEDNKNLSLYLQI 372  
QY 492 HRIKGVVTDKFGKPKVNARISVKGIHDIITAPDGDYWRLLPPGHIHIVIAQAPGAKVI 551  
Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGYWRLLVPGNYKLTASAPGYLAIT 432

Qy 461 ELGCVKPPPEALYTLQHKNKESLLNFVETVHRGKGVVTKFGKVPKNARISVKGIRHD 520  
 Db 314 ELSCKEPPPETLKNYWNEDNKNISYIQIHRGVKGFVRDLQGNPIANATLSVEGIDHD 373  
 Qy 521 ITTAPDGDYWRLLPPGHIHIAQAPGVAKVKKVIIIPARMKRAGRVDFILO 571  
 Db 374 VTSAKOGDYWRLLVPGNYKLTASAPGYLAIKAVP--YSPAVRVDFELE 422

RESULT 5  
 A54324  
 carboxypeptidase H - American goosfish  
 C:Species: Lophius americanus (American goosefish)  
 C>Date: 29-Aug-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C:Accession: A54324  
 R:Roth, W.W.; Macklin, R.B.; Spiess, J.; Goodman, R.H.; Noe, B.D.  
 Mol. Cell. Endocrinol. 78, 171-178, 1991  
 A:Title: Primary structure and tissue distribution of anglerfish carboxypeptidase  
 A:Reference number: A54324; MUID:92137483; PMID:1778303  
 A:Accession: A54324  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-454 <ROT>  
 A:Cross-references: UNIPROT:P37892; GB:S80565; NID:g244402; PIDN:AAA03252.1; P37892  
 A>Note: sequence extracted from NCBI Backbone (NCBIN:80565, NCBI:P:80566)  
 C:Superfamily: human carboxypeptidase H

Query Match 29.7%; Score 1035; DB 2; Length 454;  
 Best\_Local\_Similarity 50.1%; Pred. No. 4.2e-74;  
 Matches 207; Conservative 65; Mismatches 121; Indels 20; Gaps 7

Qy 172 IRFSHSAQMRVRLRTASCAHVARTSYIGRFDGRELIVIEFSSRPGOHELMPEVK 231  
 Db 27 ISFEYHYEELRKALVSWLQCPTRIARTYITIGSEFGEELLVLEMSDNPGETHEPGEPEFK 86  
 Qy 232 LIGNHGNVAGREMLYLAQYLCEYLLGNPRQRLNTRIHLLPSINPDGYEVAAR 291  
 Db 87 YIANMHGNEAVGRELLIYLAQYLCNQOQNETIDLTHSTRIHLMPSNPDGFEKAASQ 146  
 Qy 292 GAGYNGWTSGRQNAQLDNRFPDLTSEYVRLAETRCARSDHI-----PIPOHYWVGK 345  
 Db 147 PGETKDFVGRSNAQGVLDNRFPDLRLIY--TNEREGANNHLLQNMKKAVDEN---TK 202  
 Qy 346 VAPETKAIMKMWQITPFLVSLASLHGGDLVSPYDFSKHPQBEKMFSPDPDEKMFKLRS 405  
 Db 203 LAPETKAVIHMEIPFVLISANLHGGDVVANYPYDETGTSTHE--YSASPDVDVIFKSLAK 261  
 Qy 406 AYADVHPMMDRSENRC-----GNFLKRGSIINGADWYFTGCGSDRNYLHNTNCEFIYV 460  
 Db 262 AFSIYNPVMSPQRPCKRKHDDSSF--KDGITNGAWYSPVGGQDFNYLSSNCFEITL 319  
 Qy 461 ELGCVKPPPEALYTLQHKNKESLLNFVETVHRGKGVVTKFGKVPKNARISVKGIRHD 520  
 Db 320 ELSCDKFPNEDLTKEYTEQNRNSLVNYLEQVHRGVKGVVRLDQGNPIFNTATLSVEGIDHD 379  
 Qy 521 ITTAPDGDYWRLLPPGHIHIAQAPGVAKVKKVIIIPARMKRAGRVDFILOPL 573  
 Db 380 ITTAKGDYWRLLRQGNKYVAASAPGYLTVIKKAVP--HSPATRVDFELES 430

RESULT 6  
 S16383  
 carboxypeptidase E (EC 3.4.17.10) - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
 C:Accession: S16383  
 R:Parkinson, D.  
 submitted to the EMBL Data Library, August 1991  
 A:Description: Mutational analysis of the sorting signal for carboxypeptidase  
 A:Reference number: S16383  
 A:Accession: S16383  
 A:Molecule type: mRNA

A;Residues: 1-477 <PAR>  
A;Cross-references: UNIPROT:Q00493; EMBL:X61232; NID:g50312; PIDN:CAA43550.1; PID:g50313  
C;Superfamily: human carboxypeptidase H  
C;Keywords: hydrolase; metallo-carboxypeptidase

Query Match 29.3%; Score 1019; DB 2; Length 477;  
Best Local Similarity 44.1%; Pred. No. 8.3e-73;  
Matches 217; Conservative 69; Mismatches 142; Indels 64; Gaps 12;

QY 91 CAVLAPRCGWRPCRHICGLREVQCQAFADWMPYFDLCHRYFTFREDEGCYDPL 150  
DB 13 CAALV---AGWLL-----TREAQEPGAPAGNR-----RRRLQQDGDG----- 48  
QY 151 EKLGGLEADLPSGLPPTFIRFSSHYSYAQWVRLRRRTASRCAHVARTYSIGRSPDGE 210  
DB 49 -----ISFEHRYPELREALVSWLQCTAISRIYTVGASPEGRE 87  
QY 211 LLVIEFSSRPGQHELMPEVKLIGNHNEVAGREMLIYLAQYLCSEYLLGNPRIQLLN 270  
DB 88 LLVIELSDNPGVHEPGEPEPKYIGNHMGNEAVGRELLIFLAQYLCNEYQKGNETIVNLH 147  
QY 271 TTRIHLPSINPGYEVAAEGAGYNGWTSGRQNAQNLDLRNFPDLTSEYRLAETRG 330  
DB 148 STRIHMPSLNPGFEKAWQPGQLKDWFFVGRSNAQGLDLNRNFPDLDRIVY-VNEKEGG 206  
QY 331 RSDHI-----PIPQHYWVGKAVPETKAIMKWMOTIPFVLSASLHGDLVSVYPDFSKH 384  
DB 207 PNNHLKXKLVQDN---SKLAPETKAVIHWIMDIPFVLSANLHGDLVANTPYDETR 263  
QY 385 PQEKMFSPTPDEKFKLISRAYADVHPMMDSNRCCGN-----FLKRGSIINGADWYS 440  
DB 264 GTAHE-YSSCPDAIFQSLARAYSFPVMSDNPDPCCCKNDSDSFDGDT-TNGGAWYS 321  
QY 441 FTGMSDENYL-HTNCFEITVELGCVKFPPEEALYTLWQHNKESLNFVTHRGKGVV 499  
DB 322 VPGMGQDNVYLSNSGCFEITVELTCFPPETLKSWNEDNKNLSNYLEQIHRGVKGV 381  
QY 500 TDRFGKVPKNARISVKGIRHDIITAPDGDYWRLLPPIGHIHVAQAPYAKVKKVIPAR 559  
DB 382 RDLQGNPNANATISVDGIDHDVTSKADGDYWRLLAPGNKYLTSADFGYLAIKKVAVP-- 439  
QY 560 MKGAGRVDFILQ 571  
DB 440 FSPAVGVDFELE 451

RESULT 7  
S02074  
lysine carboxypeptidase (EC 3.4.17.3) small chain precursor - human  
N;Alternate names: anaphylatoxin inactivator small subunit; carboxypeptidase N small sub  
C;Species: Homo sapiens (man)  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 09-Jul-2004  
C;Accession: S02074  
R;Gebhard, W.; Schube, M.; Eulitz, M.  
Eur. J. Biochem. 178, 603-607, 1989  
A;Title: cDNA cloning and complete primary structure of the small, active subunit of hum  
A;Reference number: S02074; MUID:89107181; PMID:2912725  
A;Accession: S02074  
A;Molecule type: mRNA  
A;Residues: 1-458 <GEB>  
A;Cross-references: UNIPROT:P15169; EMBL:X14329; NID:g30296; PIDN:CAA32507.1; PID:g30297  
A;Note: part of this sequence, including the amino end of the mature protein, was confir  
C;Superfamily: human carboxypeptidase H  
C;Keywords: hydrolase; metallo-carboxypeptidase  
F;1-20/Domain: signal sequence #status predicted <SIG>  
F;21-458/Product: lysine (arginine) carboxypeptidase small chain #status experimental <M

Query Match 28.8%; Score 1001; DB 2; Length 458;  
Best Local Similarity 47.8%; Pred. No. 2.1e-71;  
Matches 204; Conservative 67; Mismatches 132; Indels 24; Gaps 10;

QY 172 IRFSSHYSYAQWVRLRRRTASRCAHVARTYSIGRSPDRELLVIEFSSRPGQHELMPEVK 231

DB 21 VTFRRHRYDDLVRTLVKQVQCEPGITRVVYSIGRSVEGRHLYLVLEFSDHPGIHEPLEPEVK 80  
QY 232 LIGNTHGNVAGREMLIYLAQYLCSEYLLGNPRIQLRLLNTRIHLPSINPDGYEVAAAE 291  
DB 81 YVGNMHGNEALGRELMLQSEFLCBFRNRNORIVQLIQDTRIHLPSMNPDPGYEVAAQA 140  
QY 292 GAGYNGWTSGRQNAQNLDLRNFPDL-TSEYRLAETRGASDHIPIPOHYWVGKVAPE 350  
DB 141 GPNKPGYLVGRNANGVDLNRNFPDLNTIYYN---EKYGGPNHHLPLPDN-WKQSEPE 197  
QY 351 KAIMKWMOTIPFVLSASLHGDLVSVYPDFSKHPQ-----EKMFSPTPDEKMFKLLSRA 406  
DB 198 RAVIRWHSFNFVLSANLHGAVANYPYDKSFEHVRGVRRTASTPTDDKLFOKLAKV 257  
QY 407 YADVHPMMDSNRCCGNFLKRGSIINGADWYSFTGGMSDFNYLHNTNCFEITVELGCVK 466  
DB 258 YSYAHGWMFQGW--NCGDYF--PDGITNGASWYSLSKGMQDFNYLHNTNCFEITVELSCDK 313  
QY 467 FPEEALYTLWQHNKESLNFVTHRGKGVVTDKFGKPKVNARISVKGIRHDIITAPD 526  
DB 314 FPEEELQREWIGNREALIQFLEQVHQGIKGMVLDENYNNLANAVISVSGINHDTVSGDH 373  
QY 527 GDYWRLLPPIGHIHVAQAPGY-AKVIKKVIIIPARMKACRGRVDFILQ-----PLGMGPK 578  
DB 374 GDYFRLLPPIGIVTSATAPGYDPETVTVGFA---EPTLVNFHLKRSIPQVSPVRRAPS 430  
QY 579 NFIHGLR 585  
DB 431 R-RHGV 436

RESULT 8  
S51739  
transcription repressor ABBP1 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 09-Jul-2004  
C;Accession: S60227; S51739  
R;He, G.P.; Muise, A.; Li, A.W.; Ro, H.S.  
Nature 378, 92-96, 1995  
A;Title: A eukaryotic transcriptional repressor with carboxypeptidase activity.  
A;Reference number: S60227; MUID:96061010; PMID:7477299  
A;Accession: S60227  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-719 <HE2>  
A;Cross-references: UNIPROT:Q61281; EMBL:X80478; NID:g607131; PIDN:CAA56648.1; PID:g6071

Query Match 26.0%; Score 904.5; DB 2; Length 719;  
Best Local Similarity 38.2%; Pred. No. 1.8e-63;  
Matches 198; Conservative 73; Mismatches 165; Indels 83; Gaps 11;

QY 174 FSHHYSYAQWVRLRRRTASRCAHVARTYSIGRSPDRELLVIEFSSRPGQHELMPEVKUI 233  
DB 144 FRHSHYKDMRQLMKAVNEECPTITRTYSLGSSRGLKIYAMEISDNPQDHELGEPEFRYT 203  
QY 234 GNIHNEVAGREMLIYLAQYLCSEYLLGNPRIQLRLLNTRIHLPSINPDGYEVAAAECA 293  
DB 204 AGIHNEVLGRELMLLQYLCQETRDGNPRVRNVLVQDTRIHLVPSLNPDPGYEVAAQMS 263  
QY 294 GYNGWTSGRQNAQNLDLRNFPDLTSEYRYLAETR---GARSDDHIPIQHYW--GKVA 347  
DB 264 EFGNWALGLWTEEGDFIDFEDPDNLNSVLWAAEEKKVVPRVFNPNLPIPERYLSPDATVS 323  
QY 348 PETKAIMKWMOTIPVLSASLHGDLVSVYPDFSKHPQHELMPEVK 392  
DB 324 TEVRAIIISWMEKNPVLGANLNGSERLVSYPDMARTSQEQLLAELAAARAGEDDGV 383  
QY 393 ---PTPDEKMFKLLISRAYADVHPMMDSNRCCGNFLKRG-SIINGADWYSFTGMSDF 448  
DB 384 EAQETPDHAIKFWLALISASAHLTWTEPYRGCCQADYTSGMGIVNGAKNPRSTFND 443  
QY 449 NYLHNTNCFEITVELGCVKFPPEEALYTLWQHNKESLNFVTHRGKGVVTDKFGKPKV 508

Db 444 SYLHTNCLSESVYLGCDKPFHESELPREWENKEALLTFMEQVHRGKGVVTDQGIPIA 503  
Qy 509 NARISVKGIRHDIITAPDGYWRLPPGHIHVIQAQPGYAKVKKVVIIPARMKRAGRVDF 568  
Db 504 NATISVSGINHGKVTASGGDYWRILNPGEVYRVTAAHAGYTSSAK-----ICNVYD 553  
Qy 569 ILQPLGMPKXNFI-----HGLR---RFGPHDPLGGA----- 596  
Db 554 ---DIGATQCNFI--LARSNWKRIREILAMNGRPIRLVDRSPMTPOQRRMQORRLOVRLR 610  
Qy 597 -----SSLGATEPDP-LRARRQPS-ADGSKPW 622  
Db 611 MREQMRLRLNLTAGPATSTPMPALPPSPPTAITLRPW 649

RESULT 9  
JC5256  
adipocyte transcription factor, ABBP1 - human  
C/Species: Homo sapiens (man)  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C/Accession: JC5256  
R;Ohno, I.; Hashimoto, J.; Shimizu, K.; Takaoka, K.; Ochi, T.; Matsubara, K.; Okubo, K.  
Biochem. Biophys. Res. Commun. 228, 411-414, 1996  
A/Title: A cDNA cloning of human ABBP1 from primary cultured osteoblasts and its expres  
A/Reference number: JC5256; MUID: 97079196; PMID: 8920928  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-845 <OHN>  
A/Cross-references: UNIPROT:Q14113; DDBJ:D86479; NID:g1458942; PIDN:BAAL3094.1; PID:g146

Query Match 25.7%; Score 895.5; DB 2; Length 845;  
Best Local Similarity 37.3%; Pred. No. 1.2e-62;  
Matches 200; Conservative 74; Mismatches 161; Indels 101; Gaps 12;

Qy 174 FSHSYAQMVRLRTRASCAHVARTYSIGRSPDRELLLVIEFSRPGQHELMPEVKLI 233  
Db 249 FRHSYKDMQMLKVNNEECPTTTRYSLGKSRGLKIYAMEISDNPEHELGEPEFRT 308  
Qy 234 GNTHGNVAGREMLIYLAQVLCSEYLLGNPRIQLRLNTTRHLLPSINPDGYEVAAGA 293  
Db 309 AGHNGNEVLGRELLELLLMQYLCEYRDGNRPVRSLSVQDTRHLLVPSINPDGYEVAAGMS 368  
Qy 294 GYNGWTSGRQNAQLDLNRNFPDLTSEYYRLAETR-----GARSHPHPIPHYWW--GKVA 347  
Db 369 EFGNWLGLWTEBGFDFEDFDLNSVLWGAESEKWPYRVVNNLPPIPERYLSPDATVS 428  
Qy 348 PETKAIMKWMQTTIPFVLASLHGGDLVVSYPDFSKHPQBEKMPF----- 392  
Db 429 TEVRAIIAMKEKNPFVLGAMLINGERLVSYPDYDMARTPTQQLLAAMAAARGEDEDEVS 498  
Qy 393 ---PTPDEKMFKLLSRAYADVHPMMDRSENRCGNFLKEG-SIINGADWYSFTGMSDF 448  
Db 489 EAQETPDHAIIRWLATISFASAHLTLTPEYRGCGQAQDYTCGMGIVNGAKWNPRTGINDF 548  
Qy 449 NYLHTNCFEITVELGCVKFPPEEALYTLWQHNKESLNFVETVHRGKGVVTDKFGKPVK 508  
Db 549 SYLHTNCLSESVYLGCDKPFHESELPREWENKEALLTFMEQVHRGKGVVTDQGIPIA 608  
Qy 509 NARISVKGIRHDIITAPDGYWRLPPGHIHVIQAQPGYAKVKKVVIIPARMKRAGRVDF 568  
Db 609 NATISVSGINHGKVTASGGDYWRILNPGEVYRVTAAHAGYTSPAKTC-----NVYD 658  
Qy 569 ILQPLGMPKXNFIHGLRGTG-----PH-DP----- 592  
Db 659 ---DIGATQCNFI--LARSNWKRIREIMAMNGRPIPHIDFSPMTPOQRRMQORRLOVRLR 713  
Qy 593 -----LGGASSIGATEPDP-LRARRQPSADGSKPWWSYFTSLSTHTRPW 637  
Db 714 LRLEAQMRLRLNLTATTLTGHTVPTTL-----PPAPA-----TTLSTTIEPW 755

RESULT 10

150090  
carboxypeptidase gp180 - Anas sp.  
C/Species: Anas sp.  
C/Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 09-Jul-2004  
C/Accession: I50090  
R;Kuroki, K.; Eng, F.; Ishikawa, T.; Turck, C.; Harada, F.; Ganem, D.  
J. Biol. Chem. 270, 15022-15028, 1995  
A/Title: gp180, a host cell glycoprotein that binds duck hepatitis B virus particles, i  
A/Reference number: A57010; MUID: 95318059; PMID: 7797483  
A/Accession: I50090  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1389 <KUR>  
A/Cross-references: UNIPROT:Q90240; EMBL:U25126; NID:g1008477; PIDN:AAA78903.1; PID:g10

Query Match 24.7%; Score 861; DB 2; Length 1389;  
Best Local Similarity 43.4%; Pred. No. 1.3e-59;  
Matches 188; Conservative 67; Mismatches 126; Indels 52; Gaps 12;

Qy 159 ADEALPSGLPPTP-----IRFSHSHSYAQMVRLRTRASCAHVARTYS 201  
Db 478 ATPAPPSTLTPSVAQVEPPATTSLHQAVQVDPFRHHFSDMEIFLRRYANEYPSITRLYS 537  
Qy 202 IGRSPDRELLLVIEFSRPGQHELMPEVKLIIGNIHGNEVAGREMLIYLAQVLCSEVLLG 261  
Db 538 VGKSVRELRYLMEISDNFPIHEAGEPEFKYIGNHMEHNEVVGRELLLNLTLEYLCKNPF--G 595  
Qy 262 -NPRIQLRLNTTRHLLPSINPDGYEVAAGAEGYNGWTSGRQNAQLDLNRNFPDLTSE 320  
Db 596 TDEVTDLVQSTRIHIMPENPDGYEKSQ---EGDRGGTVGRNNSNNYDLNRNFPD---Q 649  
Qy 321 YRLAETRGARSDHIPPIQHYMWGKVAPETKAIMKWMQTTIPFVLASLHGGDLVVSYPFD 380  
Db 650 FQGVTPD-----PQ-----PETLAVMSMLKTYVPFVLSANLHGGSLVNVNYPFD 691  
Qy 381 FSKHPQEEKWFTPTDEKMFKLLSRAYADVHPMMDRS---ENRCGNFLKRGSIINGADW 438  
Db 692 DDE--QGIAIYKSPDPAVFQQLALSYSENKQYQGSCKDLYPTEYFPHG-ITNGAQW 748  
Qy 439 YSTFGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHNKESLNFVETVHRGKGV 498  
Db 749 YNVPGGQDMNYLNTNCFEITVELGCVKFPKABELPKYWEQNRRLSLQFIKQVHRGIWGF 808  
Qy 499 VTDKPF-GKPVKNARISVKGIRHDIITAPDGYWRLPPGHIHVIQAQPGYAKVKKVVIIP 557  
Db 809 VLDTDGRGLNATISVADINHPVTYKGGDYWRLLVQGTYKYVTASARGYDPTVKTEVVD 868  
Qy 558 ARMKRAGRVDPIFIL 570  
Db 869 S--KGGVQVNFIL 879

RESULT 11  
T30916  
carboxypeptidase D (EC 3.4.17.-) - California sea hare  
C/Species: Aplysia californica (California sea hare)  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30916  
R;Fan, X.; Qian, Y.; Fricker, L.D.; Akalal, D.B.; Nagle, G.T.  
DNA Cell Biol. 18, 121-132, 1999  
A/Title: Cloning and expression of Aplysia carboxypeptidase D, a candidate prohormone-p  
A/Reference number: Z20933; MUID: 99171579; PMID: 10073571  
A/Accession: T30916  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-1446 <FAN>  
A/Cross-references: UNIPROT:O77063; EMBL:AF007570; NID:g3642735; PID:g3642736; PIDN:AAAC  
C/Keywords: hydrolase; metallo-carboxypeptidase

Query Match 23.7%; Score 823.5; DB 2; Length 1446;  
Best Local Similarity 39.8%; Pred. No. 1.3e-56;  
Matches 173; Conservative 78; Mismatches 137; Indels 47; Gaps 11;



## Tl3420

probable carboxypeptidase (EC 3.4.17.-) - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*

C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
C:Accession: Tl3420

R:Madueno, E.; de Pablos, B.; Modolelli, J. 1999

submitted to the EMBL Data Library, April 1999  
A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.

A:Reference number: Z17659

A:Accession: Tl3420

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1404 <NAD>

A:Cross-references: UNIPROT:P42787; EMBL:AL009147; NID:e1314051; PID:e1248570; PIDN:CAAA

A:Cross-references: FlyBase:FBgn0023535

A:Introns: 150/3; 229/2; 389/2; 423/1; 462/2; 904/3; 979/2

A>Note: EG:171D11.3

C:Keywords: hydrolase; metallo-carboxypeptidase

Query Match 21.6%; Score 750; DB 2; Length 1404;  
Best Local Similarity 39.9%; Pred. No. 8.4e-51;  
Matches 172; Conservative 71; Mismatches 132; Indels 56; Gaps 15;

QY 196 VARTYSTGRSGDRELIVIFSSRPQGH---ELMEPEVKLIHNEVAGREMLIYLAQ 252

DB 55 LAQTYTIGKSLDRPIYALALSAPTGESKNGDLLRPWVKLVANIQQDEAVGRQWLYMAE 114

QY 253 YLSEYLLGNPRQRLNLTTRIHLPSINPDGYEVAAG-----AGYNGWTSGRQNAQN 307

DB 115 YLATHY-DGDPKQALLNLTEIHLPTCPDGP-AKAKEGNCESLPNY-----VGRGNAAN 168

QY 308 LDLRNPPDL--TSEYVRL-AETRGARSDHIPQHYWGWKVPAPETKAIKMWQITPFPVL 364

DB 169 IDLRNPPDLRLEQSHVHQLRAQR-----QPETAALVNVIVSRFPVL 210

QY 365 SASLHGDLVVSYPFDFSKHPQEEKMFSPTDEKMFKLLSRAYADVHPMMDRSENRCGG 424

DB 211 SANFHGGAVVASYPYDNSLAHNECCESLTPDDRVPKQLAHTYSDNHPIM--RKGNNCND 268

QY 425 NFLKSGSIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKPPPEEALYTLWQHKESL 484

DB 269 SF--SGGITNGAHWYELSGGMQDFNYAFSNCFFELTIELSCCKYPAASTLTPQEWQNKASL 326

QY 485 LNFVETVHRGIGKVVDKFGKPVKNARISVKGIH--DITTA PDGYWRLLPGLHIVIAQ 543

DB 327 LQLLRQAHIGIKGLVTDASGFPADANVYVAGLEEKPMRTSKRGEYWRLLTPGLISVHAS 386

QY 544 APGY-AKVIVKVIIPARMKRAGRVDFILOPLGMGPKNFHGLRRTGPHDPLGASSIGE 602

DB 387 AFGVQTSAPQQRVTNDNQEARLDFKLPV---ETNFDGNFRKVKV-----ER 432

QY 603 TEPDPLRARRQ 613

DB 433 SEP-PQKLKKQ 442

## RESULT 15

## Tl3284

carboxypeptidase (EC 3.4.17.-) precursor - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000

C:Accession: Tl3284

R:Settle, S.H.; Green, M.M.; Burtis, K.C.

Proc. Natl. Acad. Sci. U.S.A. 92, 9470-9474, 1995

A>Title: The silver gene of *Drosophila melanogaster* encodes multiple carboxypeptidases

A:Reference number: Z17649; MUID:96003800; PMID:7568156

A:Accession: Tl3284

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1119 <SET>

A:Cross-references: EMBL:U29591; NID:g974552; PID:g974553; PIDN:AAA91650.1

C:Genetics:

A:Gene: svr

A:Cross-references: FlyBase:FBgn0004648

A:Map position: X

C:Keywords: hydrolase; metallo-carboxypeptidase

Query Match 21.4%; Score 743.5; DB 2; Length 1119;  
Best Local Similarity 36.3%; Pred. No. 2e-50;  
Matches 180; Conservative 74; Mismatches 151; Indels 91; Gaps 15;

QY 135 CHRYFTREDS-----GCYDPLEKLGLEADEALPSGLPTFFIRFSSHSAQMVRVL 186

DB 23 CROYTIKEDSFLOQPHYASQEQLEDLFAGLE--KAYP----- 58

QY 187 RRTASRCARHARTYSTIGRSDFRELIVIFSSRPQGH---ELMEPEVKLIHNEVAGREML 246

DB 59 -----NQAKVHFLGRSLRNLALQISRTRSRNLLTPVKYIANHMGDETVCRL 110

QY 247 LIYLAQVLCSEYLLGNPR-----TORLNTTRIHLPSINPDGYEVAAGAGYNGWTSGR 302

DB 111 LVYMAQ-----YLLGNHERISDLGQLVNSTDIYLVPTMNPDPYALSOEGNCESLPNYVGR 165

QY 303 QNAQNLDLRNPPDL--TSEYVRL-AETRGARSDHIPQHYWGWKVPAPETKAIKMWQIT 359

DB 166 GNAANIDLNRDPPDLRLEQSHVHQLRAQR-----QPETAALVNVIVS 207

QY 360 IPFVLSASLHGDLVVSYPFDFSKHPQEEKMFSPTDEKMFKLLSRAYADVHPMMDRSE 419

DB 208 KPFVLSANFHGGAVVASYPYDNSLAHNECCESLTPDDRVPKQLAHTYSDNHPIM--RKG 265

QY 420 NRCGGNFKRGSIIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKPPPEEALYTLWQH 479

DB 266 NNCNDSF--SGGITNGAHWYELSGGMQDFNYAFSNCFFELTIELSCCKYPAASTLTPQEWQ 323

QY 480 NKESILNFTVETVHRGIGKVVDKFGKPVKNARISVKGIH--DITTA PDGYWRLLPGLH 538

DB 324 NKASLLQLLRQAHIGIKGLVTDASGFPADANVYVAGLEEKPMRTSKRGEYWRLLTPGL 383

QY 539 IVIAQAPGY-AKVIVKVIIPARMKRAGRVDFILOPLGMGPKNFHGLRRTGPHDPLGAS 597

DB 384 SVHASAFGVQTSAPQQRVTNDNQEARLDFKLPV---ETNFDGNFRKVKV----- 432

QY 598 SLGEATEPDPPLRARRQ 613

DB 433 ---ERSEP-PQKLKKQ 444

Search completed: January 25, 2005, 08:53:11

Job time : 45 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 25, 2005, 08:34:29 ; Search time 157 Seconds  
(without alignments)  
1464.621 Million cell updates/sec

Title: US-10-757-262-104

Perfect score: 3480

Sequence: 1 MPPPPLLLLLVVAAARP.....WWSYFTSLSTRPRLKY 641

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: A\_Geneseq\_23Sep04:\*
- 2: Geneseqp1980s:\*
- 3: Geneseqp1990s:\*
- 4: Geneseqp2000s:\*
- 5: Geneseqp2001s:\*
- 6: Geneseqp2002s:\*
- 7: Geneseqp2003as:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3480	100.0	641	3	AAY91044 Human car
2	3480	100.0	641	8	ADQ18097 Human sof
3	1849.5	53.1	364	6	ADA54913 Human pro
4	1058	30.4	476	5	Abb78286 Amino aci
5	1057	30.4	476	3	AAY91037 Rat carbo
6	1057	30.4	476	7	ADe83518 Rat Prote
7	1057	30.4	476	7	ADDA46262 Rat Prote
8	1054	30.3	476	3	AAY91036 Human car
9	1054	30.3	476	4	AAM23880 Human EST
10	1054	30.3	476	5	ABG61920 Prostate
11	1054	30.3	476	5	AAU84326 Protein C
12	1054	30.3	476	5	Abb78283 Amino aci
13	1054	30.3	476	7	ADDA46264 Human pro
14	1054	30.3	476	7	ADN39258 Cancer/an
15	1054	30.3	476	8	ADQ17903 Human sof
16	1052	30.2	434	3	AAY91041 Alysia ca
17	1052	30.2	434	5	Abb78288 Amino aci
18	1051	30.2	491	3	AA57137 Human pro
19	1051	30.2	491	4	AAG73399 Human gen
20	1051	30.2	491	4	AAE04143 Human gen
21	1051	30.2	491	5	ABG64310 Human alb
22	1051	30.2	491	5	ABG64288 Human alb
23	1051	30.2	491	8	ADL77553 Albumin f
24	1051	30.2	491	8	ADL77575 Albumin f
25	1049	30.1	476	3	AAY91040 Murine ca

26	1049	30.1	476	5	ABB78287	Abb78287 Amino aci
27	1037	29.8	734	8	ADN33957	Adn33957 Human nov
28	1036	29.8	714	8	ADO42309	Ado42309 Human NOV
29	1036	29.8	720	8	ADO42313	Ado42313 Human NOV
30	1036	29.8	734	4	ABA47184	Ab47184 ACPLX pro
31	1036	29.8	734	4	ABA74694	Ab74694 Human pro
32	1036	29.8	734	4	AAU29252	Aau29252 Human PRO
33	1036	29.8	734	4	AAG65917	Ag65917 Amino aci
34	1036	29.8	734	6	ABU58628	Abu58628 Human PRO
35	1036	29.8	734	6	ABU88176	Abu88176 Novel hum
36	1036	29.8	734	6	ABU84491	Abu84491 Human sec
37	1036	29.8	734	6	ABR66365	Ab66365 Human sec
38	1036	29.8	734	6	ABR65755	Ab65755 Human sec
39	1036	29.8	734	6	ABU99695	Abu99695 Human sec
40	1036	29.8	734	6	ABU82934	Abu82934 Human PRO
41	1036	29.8	734	6	ABU90055	Abu90055 Novel hum
42	1036	29.8	734	6	ABR68304	Ab68304 Human sec
43	1036	29.8	734	6	ABU96357	Abu96357 Novel hum
44	1036	29.8	734	6	ABU92788	Abu92788 Human sec
45	1036	29.8	734	6	ABO08865	Ab08865 Human sec

ALIGNMENTS

RESULT 1

AAAY91044

ID AAY91044 standard; protein; 641 AA.

AC AAY91044;

XX

DT 07-SEP-2000 (first entry)

XX

DE Human carboxypeptidase homologue CP2 protein sequence SEQ ID NO:10.

XX

XX Carboxypeptidase E; CPE; detection; diagnosis; mutation;

KW type II diabetes; antidiabetic.

XX

OS Homo sapiens.

XX

PN WO200023784-A2.

XX

PD 27-APR-2000.

XX

PF 21-OCT-1999; 99WO-US024566.

XX

PR 21-OCT-1998; 98US-0105102P.

PR 19-JAN-1999; 99US-00233989.

XX

(MILL-) MILLENNIUM PHARM INC.

Chen H, Meyer J;

XX

WPI; 2000-339757/29.

Diagnosing the presence of and/or predisposition to type II diabetes by detecting alterations in expression, sequence or function of carboxypeptidase E gene or gene product obtained from a biological sample of a subject.

Disclosure; Page 74-77; 77pp; English.

The present invention describes a method for detecting a mutation in the carboxypeptidase E (CPE) gene, or gene product, in a subject having, or predisposed to having type II diabetes. The method comprises obtaining a sample from a subject, where the sample contains a CPE gene, or gene product, and detecting an alteration in the CPE gene or gene product. CPE nucleic acids and peptides can be used to increase CPE levels in the subject. CPE antibodies can be used to modulate CPE levels. The method can be used for diagnosing, treating and preventing type II diabetes. CPE polypeptides and polynucleotides are useful for treating and diagnosing type II diabetes. They are also useful for identifying compounds that can modulate expression or function of CPE, and are useful for treatment and

CC diagnosis of Type II diabetes. The present sequence represents a human  
 CC carboxypeptidase homologue protein sequence, which is used in the  
 CC exemplification of the present invention  
 XX  
 SQ Sequence 641 AA;

Query Match 100.0%; Score 3480; DB 3; Length 641;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPPPPPLLLLTVLVVAARPGCEFERNPAATCVDLQRTCSDAAYNHTTFFNLLQHRWE 60  
 DB 1 MPPPPPLLLLTVLVVAARPGCEFERNPAATCVDLQRTCSDAAYNHTTFFNLLQHRWE 60

QY 61 VVEASSEYILLSVLHQLLEGQCNPDRLGCAVLAPRCEGGWVRRRCRHCIEGLREVCP 120  
 DB 61 VVEASSEYILLSVLHQLLEGQCNPDRLGCAVLAPRCEGGWVRRRCRHCIEGLREVCP 120

QY 121 AFDAIDMAWPFYLDCHRYFTREDEGCVDPLEKLRGGLDEALPSGLPTTFRFSSHVA 180  
 DB 121 AFDAIDMAWPFYLDCHRYFTREDEGCVDPLEKLRGGLDEALPSGLPTTFRFSSHVA 180

QY 181 QMVRVLRRTASRCARHARTYSIGRSFDRGELLVIEFSRPGQHELMPEVKLIGNHNE 240  
 DB 181 QMVRVLRRTASRCARHARTYSIGRSFDRGELLVIEFSRPGQHELMPEVKLIGNHNE 240

QY 241 VAGREMLIYLAQYLCSEYLLGNPRIORLLNTTTHILLPSINPDGYVAAAGAGYNGWTS 300  
 DB 241 VAGREMLIYLAQYLCSEYLLGNPRIORLLNTTTHILLPSINPDGYVAAAGAGYNGWTS 300

QY 301 GRQNAQNLDLNRNFPDLTSEYRLAETRGARSDHIPQHYWVGKVPAPETKAIKMWQTI 360  
 DB 301 GRQNAQNLDLNRNFPDLTSEYRLAETRGARSDHIPQHYWVGKVPAPETKAIKMWQTI 360

QY 361 PFVLSASLHGGDLVVSYPDFSKHPQBEKMFSPDDEKMFLLSRAYADVHPMMDRSEN 420  
 DB 361 PFVLSASLHGGDLVVSYPDFSKHPQBEKMFSPDDEKMFLLSRAYADVHPMMDRSEN 420

QY 421 RCGGNFLKRGSIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHN 480  
 DB 421 RCGGNFLKRGSIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHN 480

QY 481 KESLLNFVETVHRGIKGVVTDKFGKPVKNARISVKGIRHDITITAPDGDYWRLLPPGIHIV 540  
 DB 481 KESLLNFVETVHRGIKGVVTDKFGKPVKNARISVKGIRHDITITAPDGDYWRLLPPGIHIV 540

QY 541 IQAQPAGYAKVIKKVIIIPARKKAGRVDFILOPLGMPKPNFIHGLRRTGPHDPLGGASSLG 600  
 DB 541 IQAQPAGYAKVIKKVIIIPARKKAGRVDFILOPLGMPKPNFIHGLRRTGPHDPLGGASSLG 600

QY 601 EATEPDLARRQPSADGSKPWWWSYFTSLSTRPRWLLKY 641  
 DB 601 EATEPDLARRQPSADGSKPWWWSYFTSLSTRPRWLLKY 641

## RESULT 2

ADQ18097

ID ADQ18097 standard; protein; 641 AA.

XX AC ADQ18097;

XX AC ADQ18097;

DT 26-AUG-2004 (first entry)

DE Human soft tissue sarcoma-upregulated protein - SEQ ID 914.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX Homo sapiens.

XX WO2004048938-A2.

XX 10-JUN-2004.

XX

PF 26-NOV-2003; 2003WO-US038193.  
 XX  
 PR 26-NOV-2002; 2002US-0429739P.  
 XX  
 PA (PROT-) PROTEIN DESIGN LABS INC.  
 XX  
 PI Aziz N, Ginsburg WM, Zlotnik A;  
 XX  
 DR WPI; 2004-441208/41.

XX Early detection of soft tissue sarcoma comprises determining expression  
 PT of a gene in a first soft tissue sample and a normal soft tissue sample  
 PT and comparing the gene expression, also useful in treating soft tissue  
 PT sarcoma.

PS Example 2; SEQ ID NO 914; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma  
 CC which comprises obtaining a first soft tissue sample from an individual  
 CC and a normal soft tissue sample from the same or different individual,  
 CC determining the expression of a gene in both samples and comparing the  
 CC expression of the gene in both soft tissue samples, where a higher level  
 CC of protein expression in the first soft tissue sample indicates the  
 CC presence of soft tissue sarcoma. The method of the invention has  
 CC cytostatic applications and may be useful for detecting soft tissue  
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 CC acid sequences may be useful in diagnostic and screening applications.  
 CC The current sequence is that of a human soft tissue sarcoma-upregulated  
 CC protein of the invention. The current sequence is not shown within the  
 CC specification per se but was submitted in CD format by the inventor.

SQ Sequence 641 AA;

Query Match 100.0%; Score 3480; DB 8; Length 641;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 641; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPPPPPLLLLTVLVVAARPGCEFERNPAATCVDLQRTCSDAAYNHTTFFNLLQHRWE 60  
 DB 1 MPPPPPLLLLTVLVVAARPGCEFERNPAATCVDLQRTCSDAAYNHTTFFNLLQHRWE 60

QY 61 VVEASSEYILLSVLHQLLEGQCNPDRLGCAVLAPRCEGGWVRRRCRHCIEGLREVCP 120  
 DB 61 VVEASSEYILLSVLHQLLEGQCNPDRLGCAVLAPRCEGGWVRRRCRHCIEGLREVCP 120

QY 121 AFDAIDMAWPFYLDCHRYFTREDEGCVDPLEKLRGGLDEALPSGLPTTFRFSSHVA 180  
 DB 121 AFDAIDMAWPFYLDCHRYFTREDEGCVDPLEKLRGGLDEALPSGLPTTFRFSSHVA 180

QY 181 QMVRVLRRTASRCARHARTYSIGRSFDRGELLVIEFSRPGQHELMPEVKLIGNHNE 240  
 DB 181 QMVRVLRRTASRCARHARTYSIGRSFDRGELLVIEFSRPGQHELMPEVKLIGNHNE 240

QY 241 VAGREMLIYLAQYLCSEYLLGNPRIORLLNTTTHILLPSINPDGYVAAAGAGYNGWTS 300  
 DB 241 VAGREMLIYLAQYLCSEYLLGNPRIORLLNTTTHILLPSINPDGYVAAAGAGYNGWTS 300

QY 301 GRQNAQNLDLNRNFPDLTSEYRLAETRGARSDHIPQHYWVGKVPAPETKAIKMWQTI 360  
 DB 301 GRQNAQNLDLNRNFPDLTSEYRLAETRGARSDHIPQHYWVGKVPAPETKAIKMWQTI 360

QY 361 PFVLSASLHGGDLVVSYPDFSKHPQBEKMFSPDDEKMFLLSRAYADVHPMMDRSEN 420  
 DB 361 PFVLSASLHGGDLVVSYPDFSKHPQBEKMFSPDDEKMFLLSRAYADVHPMMDRSEN 420

QY 421 RCGGNFLKRGSIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHN 480  
 DB 421 RCGGNFLKRGSIINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLWQHN 480

QY 481 KESLLNFVETVHRGIKGVVTDKFGKPVKNARISVKGIRHDITITAPDGDYWRLLPPGIHIV 540  
 DB 481 KESLLNFVETVHRGIKGVVTDKFGKPVKNARISVKGIRHDITITAPDGDYWRLLPPGIHIV 540

QY 541 IAQAPGYAKVKKVILPARKRAGRVDFILOPLGMGPKNFIHGLRRTGPHDPLGGASSLG 600  
DB 541 IAQAPGYAKVKKVILPARKRAGRVDFILOPLGMGPKNFIHGLRRTGPHDPLGGASSLG 600  
QY 601 EATEPDLRRAROPSADGSKPWWSYFTSLSTRPRLKX 641  
DB 601 EATEPDLRRAROPSADGSKPWWSYFTSLSTRPRLKX 641

RESULT 3  
ADA54913  
ID ADA54913 standard; protein; 364 AA.  
XX  
AC ADA54913;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Human protein, SEQ ID 2481.  
XX  
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
Gene Therapy; human; secretory protein; membrane proteins; cancer;  
inflammatory disease; osteoporosis; neurological disease.  
XX  
OS Homo sapiens.  
XX  
PN EP1293569-A2.  
XX  
PD 19-MAR-2003.  
XX  
PF 21-MAR-2002; 2002EP-00006586.  
XX  
PR 14-SEP-2001; 2001JP-00328381.  
PR 24-JAN-2002; 2002US-0350435P.  
XX  
PA (HELI-) HELIX RES INST.  
PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX  
DR WPI; 2003-395539/38.  
DR N-PSDB; ADA53274.  
XX  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
PS Claim 14; SEQ ID NO 2481; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
SQ Sequence 364 AA;

Query Match 53.1%; Score 1849.5; DB 6; Length 364;  
Best Local Similarity 96.4%; Pred. No. 1.9e-172;  
Matches 344; Conservative 1; Mismatches 1; Indels 11; Gaps 1;  
QY 1 MPPPPPLLLTLVLVAAARPGCFERNPA-----ATCVDLQLRTCSDAAYNHTT 49  
DB 1 MPPPPPLLLTLVLVAAARPGCFERNPAGECHRPAAADSATCVDLQLRTCSDAAYNHTT 60  
QY 50 FPNLLQHRSEWVEASSEYILLVHQLLEGQCNPDRLRLGCAVLAPRCGEGWVRPCRH 109  
DB 61 FPNLLQHRSEWVEASSEYILLVHQLLEGQCNPDRLRLGCAVLAPRCGEGWVRPCRH 120  
QY 110 ICGLEVCQAFADMDMWPYFLDCHRYTTRDEGCDPFLKRLGLEADRALPSGLPP 169  
DB 121 ICGLEVCQAFADMDMWPYFLDCHRYTTRDEGCDPFLKRLGLEADRALPSGLPP 180

QY 170 TFIHFHSHSYAQWVRVLRRTASCAHVARTYSIGRSFDGRELIVIFSSRPGQHELMEPE 229  
DB 181 TFIHFHSHSYAQWVRVLRRTASCAHVARTYSIGRSFDGRELIVIFSSRPGQHELMEPE 240  
QY 230 VKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIQLLNTTRIHLPLPSINPDGYVAA 289  
DB 241 VKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGNPRIQLLNTTRIHLPLPSINPDGYVAA 300  
QY 290 AEGAGYNGWTSGRQNAQNLDLNRNFPDLTSEYVRLAETRGARSDHIPPIPOHYWVGK 346  
DB 301 AEGAGYNGWTSGRQNAQNLDLNRNFPDLTSEYVRLAETRGARSDHIPPIPOHYWVGK 357

## RESULT 4

ABB78286  
ID ABB78286 standard; protein; 476 AA.  
XX  
AC ABB78286;  
XX  
DT 05-DEC-2002 (first entry)  
XX  
DE Amino acid sequence of rat carboxypeptidase E.  
XX  
KW Wolframin; carboxypeptidase E; binding partner.  
XX  
OS Rattus sp.  
XX  
PN WO200263307-A2.  
XX  
PD 15-AUG-2002.  
XX  
PF 30-JAN-2002; 2002WO-US001226.  
XX  
PR 02-FEB-2001; 2001US-0266385P.  
XX  
PA (PHAA ) PHARMACIA & UPJOHN CO.  
XX  
PI Hiebsch RR;  
XX  
DR WPI; 2002-627571/67.  
DR N-PSDB; ABV72124.  
XX  
XX Identifying agents for increasing propensity of Wolframin protein to  
PT associate with carboxypeptidase E binding partner polypeptide, by  
PT contacting proteins in presence of agent and determining increased  
PT propensity.  
XX  
PS Claim 29; Page 75-77; 85pp; English.  
XX  
CC The present sequence represents carboxypeptidase E. The specification  
CC describes a method for identifying agents which increase the propensity  
CC of wolframin protein to associate with a carboxypeptidase E binding  
CC partner polypeptide. The method comprises contacting both polypeptides in  
CC the presence of a test agent, and determining whether the association of  
CC wolframin with carboxypeptidase E is increased. Such agents can be used  
CC to modulate wolframin protein activity  
XX  
SQ Sequence 476 AA;

Query Match 30.4%; Score 1058; DB 5; Length 476;  
Best Local Similarity 48.4%; Pred. No. 1.9e-94;  
Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;  
QY 155 GGLADEALPSGLPPTFIR-----FSHHSYAQMVRVLRRTASCAHVARTYSI 202  
DB 20 GWLLAAEAQEPGAPAAAGMRRRLQOEDGISPEYHRYPELREALVSWLQCTAISRIYTV 79  
QY 203 GRSFDCRELLIVTEFSSRPGQHELMEPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGN 262  
DB 80 GRSFEGRELLIVTELSNPGVHEPFEFKYIGNMGHNEAVGRELILFLAQLYLCNEORGN 139  
QY 263 PRIORLNTTRIHLPLPSINPDGYEVAAGAGYNGWTSGRQNAQNLDLNRNFPDLTSEY 322

Db 140 ETIVNLHSTRHIMPSLNFDPGFEKASQFGLKDFVGRSNAQGDIDLNENFDDLRIVY 199  
 QY 323 RLAEATRGARDHI-----PIQHYWVGKVPAPETKAIMKWMQTIPIFVLSASLHGGDLVVS 376  
 Db 200 -VNEKEGGPNHLLKNLKKIVDQN---SKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255  
 QY 377 YPFDSEKHPQEEKMFSPDDEKMFLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
 Db 256 YPDETRSGTAHE-YSSCPDDAIFQSLARAYSFNPMVSDPNRPPCRKXNDSSVF--DG 312  
 QY 432 IINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLQHNKESLLNFVETV 491  
 Db 313 TTNGGAWYSVPGMGQDFNYLSSNCFEITVELSCEKFPPEETLKSYPEDNKNLSINYLEQI 372  
 QY 492 HRGIGKGVVTDKFGKPVKNARISVKGIHDTITAPDGDYWRLLPPGHIHIVAQAPYAKVI 551  
 Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKADGDYWRLLAPGNYKLTASAPGYLAIT 432  
 QY 552 KKVIIIPARMKRAGRVDFILQ 571  
 Db 433 KKVAVP--FSPAVGVDFELE 450

RESULT 5  
 AAY91037  
 ID AAY91037 standard; protein; 476 AA.

AC AAY91037;

DT 07-SEP-2000 (first entry)

DE Rat carboxypeptidase E protein sequence SEQ ID NO:3.

XX Carboxypeptidase E; CPE; detection; diagnosis; mutation;

KW type II diabetes; antidiabetic.

OS Rattus sp.

XX WO200023784-A2.

XX 27-APR-2000.

XX 21-OCT-1999; 99WO-US024566.

XX 21-OCT-1998; 98US-0105102P.

XX 19-JAN-1999; 99US-00233989.

XX (MILL-) MILLENNIUM PHARM INC.

XX Chen H, Meyer J;

XX WPI; 2000-339757/29.

PT Diagnosing the presence of and/or predisposition to type II diabetes by  
 PT detecting alterations in expression, sequence or function of  
 PT carboxypeptidase E gene or gene product obtained from a biological sample  
 PT of a subject.

XX Disclosure; Page 60-61; 77pp; English.

CC The present invention describes a method for detecting a mutation in the  
 CC carboxypeptidase E (CPE) gene, or gene product, in a subject having, or  
 CC predisposed to having type II diabetes. The method comprises obtaining a  
 CC sample from a subject, where the sample contains a CPE gene, or gene  
 CC product, and detecting an alteration in the CPE gene or gene product. CPE  
 CC nucleic acids and peptides can be used to increase CPE levels in the  
 CC subject, CPE antibodies can be used to modulate CPE levels. The method  
 CC can be used for diagnosing, treating and preventing type II diabetes. CPE  
 CC polypeptides and polynucleotides are useful for treating and diagnosing  
 CC type II diabetes. They are also useful for identifying compounds that can  
 CC modulate expression or function of CPE, and are useful for treatment and  
 CC diagnosis of type II diabetes. The present sequence represents rat CPE,

CC which is used in the exemplification of the present invention

XX SQ Sequence 476 AA;

Query Match 30.4%; Score 1057; DB 3; Length 476;

Best Local Similarity 48.4%; Pred. No. 2.4e-94;

Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLADEALPSGLPTTFIR-----FSHSYIAQMVRVLRTASRCAHVARTYSI 202

Db 20 GLLAAEAQEPGAPAAAGMRRRLRLQEDGISPEYHRYPELREALVSVMLQCTAISRIYTV 79

QY 203 GRSFGRELLVTEFSSRCQHELMPEVKLIGNIHGNEVAGHEMLIYLAOYLCSYLLGN 262

Db 80 GRSFGRELLVTELSNDNPGVHEPGEPEFKYIGNHMGNEAVGRELIFLAQYLCNEYQKGN 139

QY 263 PRIORILLNTRTHLPSINPDGVEVAAAEAGAGYNGWTSGRONAQLDNRNPPDLTSEY 322

Db 140 ETIVNLHSTRHIMPSLNFDPGFEKASQFGLKDFVGRSNAQGDIDLNENFDDLRIVY 199

QY 323 RLAEATRGARDHI-----PIQHYWVGKVPAPETKAIMKWMQTIPIFVLSASLHGGDLVVS 376

Db 200 -VNEKEGGPNHLLKNLKKIVDQN---SKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255

QY 377 YPFDSEKHPQEEKMFSPDDEKMFLLSRAYADVHPMMDRSENRCGN-----FLKRG 431

Db 256 YPDETRSGTAHE-YSSCPDDAIFQSLARAYSFNPMVSDPNRPPCRKXNDSSVF--DG 312

QY 432 IINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEEALYTLQHNKESLLNFVETV 491

Db 313 TTNGGAWYSVPGMGQDFNYLSSNCFEITVELSCEKFPPEETLKSYPEDNKNLSINYLEQI 372

QY 492 HRGIGKGVVTDKFGKPVKNARISVKGIHDTITAPDGDYWRLLPPGHIHIVAQAPYAKVI 551

Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKADGDYWRLLIPGNYKLTASAPGYLAIT 432

QY 552 KKVIIIPARMKRAGRVDFILQ 571

Db 433 KKVAVP--FSPAVGVDFELE 450

RESULT 6

ADE83518

ID ADE83518 standard; protein; 476 AA.

XX AC ADE83518;

XX DT 29-JAN-2004 (first entry)

XX DE Rat Protein P15087, SEQ ID NO 11115.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;

KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

OS Rattus norvegicus.

XX WO2003016475-A2.

XX PD 27-FEB-2003.

XX PF 14-AUG-2002; 2002WO-US025765.

XX PR 14-AUG-2001; 2001US-0312147P.

XX PR 01-NOV-2001; 2001US-0346382P.

XX PR 26-NOV-2001; 2001US-0333347P.

XX PA (GEO ) GEN HOSPITAL CORP.

XX PA (FARB ) BAYER AG.

XX PI Woolf C, D'urso D, Befort K, Costigan M;

XX DR WPI; 2003-268312/26.

XX DR GENBANK; P15087.

XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 476 AA;

Query Match 30.4%; Score 1057; DB 7; Length 476;

Best Local Similarity 48.4%; Pred. No. 2.4e-94;

Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADENLPSGLPPTFIR-----PSHHSYAQMVRVLRRTASCAHVARTYSI 202  
Db 20 GWLLAAEOAEPGAPAGMRRRLQEDGSGSFYHRYPELRREALVSVWLQCTAISRYTV 79  
QY 203 GRSGDRELLVIBFSSRPGOHELMPEVKLIGNHNEVAGREMLYLAQYLCSEYLLGN 262  
Db 80 GRSGFGRRELLVIBLSDNPGVHEGEPEFKYIGNHNEVAGRELLIFLAQYLCNEVQRGN 139  
QY 263 PRIORLLNTRIHLPSINPDGYEVAAGAGYNGWTSGRQNAQNLNLRNFPDLTSEY 322  
Db 140 ETIVNLHSTRHIMPSLNDPGFEKAASQPGELKDFVGRSNAQGDILNRNFPDLDRIVY 199  
QY 323 RLAEITRGARDHI-----PIPOHYWGKVAPETKAIMQMOTIPFVLSASLHGGDLVVS 376  
Db 200 -VNEKEGEGPNHLLKMLKKTVDQN---SKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255  
QY 377 YPDPFSKHQOEKMFPTPEKMFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
Db 256 YPDETSTGTAHE -YSCPDDAIFQSLARAYSFNPMDSDPNPPCKDKDDSSFF--DG 312  
QY 432 IINGADWYSFTGMSDFNYLHTNCFEITVELGCVKFPPEBALYTLQHNKESILLNFVETV 491  
Db 313 TTNGGAWYSVPGMQDFNYLSSNCFEITVELSCEKFPPEBTAKSYWEDKNKSLINYLEQI 372  
QY 492 HRGKGWVTDKFGKPVNARISVKGIRHDTTAPDGDYWRLLPPGHHIVIAQAPGVAKVI 551  
Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSAKDGDYWRLLVPGNYKLTASAPGYLAIT 432  
QY 552 KKVIIIPARMKRCRGRVDFILQ 571  
Db 433 KKVAVP--FSPAVGVDFELE 450

## RESULT 7

ADD46262

ID ADD46262 standard; protein; 476 AA.

XX AC ADD46262;

XX AC ADD46262;

DT 29-JAN-2004 (first entry)

DE Rat Protein P15087, SEQ ID NO 11937.

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
XX

OS Rattus norvegicus.

PN WO2003016475-A2.

PD 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 28-NOV-2001; 2001US-0333347P.

XX (GEO ) GEN HOSPITAL CORP.

XX (FARB ) BAYER AG.

XX Woolf C, D'urao D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; P15087.

PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

PS The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a rat protein (shown in Table 2 of  
XX the specification) which is differentially expressed during pain. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic form directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 476 AA;

Query Match 30.4%; Score 1057; DB 7; Length 476;

Best Local Similarity 48.4%; Pred. No. 2.4e-94;

Matches 213; Conservative 66; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADENLPSGLPPTFIR-----PSHHSYAQMVRVLRRTASCAHVARTYSI 202

Db 20 GLLLAQAQEPGAPAGRRRRRLRQEDGIGSFYHRYPELREALVSVWLQCTAISRIYTV 79  
 QY 203 GRSPDGRRELLVIEFSSRPGQHELMPEVVKLIIGNHNEVAGREMLIYLAQYLCSEYLLGN 262  
 Db 80 GRSPGGRRELLVIELSDNPGVHEPGEPEFKYIGNHNEVAGREMLIYLAQYLCSEYQKGN 139  
 QY 263 PRIQLRLNTTRIHLPLPSINPDGYEVAAGAGYNGWTSGRONQNLDIRNFPDLTSEY 322  
 Db 140 ETIVNLHSTRIHIMPSLNDPGFEKASQPGELKDWPFVGRSNAQGDIDLNRFPPDLDRIVY 199  
 QY 323 RLAEARGARDHI-----PIPOHYMWKGVAPETKAIMKMWOTIPFVLSASLHGGDLVVS 376  
 Db 200 -VNEKEGGPNHLLKNKIVDQN---SKLAPETKAVIHWMIDIPFVLSANLHGGDLVAN 255  
 QY 377 YPFDGSKHPOEKMFSPTPEKFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431  
 Db 256 YPYDETRSGAHE-YSSCPDDAIFQSLARAYSSFNPMSPDNRPPCRKNDSDSSFV--DG 312  
 QY 432 IINGADWYSFTGMSDFNYLHNTCFEITVELGCVKPPPEALYTLQHNKESILNFVEIV 491  
 Db 313 TTNGGAWYSVPGGMQDFNYLSSNCFEITVELSCEKFPPEETLKSYPWEDNKNLSINYLEQI 372  
 QY 492 HRGIGKVVTDKFGKPVKNARISVKGIRHDITTPADGDYWRLLPPIGHIHIVIAQAPGYAKVI 551  
 Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGWYRLLVPGNYKLTASAPGYLAIT 432  
 QY 552 KKVLIIPARMKGRVDFILQ 571  
 Db 433 KKVAVP--FSPAVGVDFELE 450

## RESULT 8

AA91036  
 ID AAY91036 standard; protein; 476 AA.

XX AC AAY91036;

XX 07-SEP-2000 (first entry)

XX Human carboxypeptidase E protein sequence SEQ ID NO:2.

XX Human; carboxypeptidase E; CPE; detection; diagnosis; mutation;  
 type II diabetes; antidiabetic.

XX Homo sapiens.

XX WO200023784-A2.

XX 27-APR-2000.

XX 21-OCT-1999; 99WO-US024566.

XX 21-OCT-1998; 98US-0105102P.

XX 19-JAN-1999; 99US-00233989.

XX (MILL-) MILLENNIUM PHARM INC.

XX Chen H, Meyer J;

XX WPI; 2000-339757/29.

XX N-PSDB; AAA39242.

XX Diagnosing the presence of and/or predisposition to type II diabetes by  
 detecting alterations in expression, sequence or function of  
 carboxypeptidase E gene or gene product obtained from a biological sample  
 of a subject.

XX Disclosure; Page 58-59; 77pp; English.

XX PS

XX The present invention describes a method for detecting a mutation in the  
 carboxypeptidase E (CPE) gene, or gene product, in a subject having, or  
 predisposed to having type II diabetes. The method comprises obtaining a

CC sample from a subject, where the sample contains a CPE gene, or gene  
 product, and detecting an alteration in the CPE gene or gene product. CPE  
 nucleic acids and peptides can be used to increase CPE levels in the  
 subject, CPE antibodies can be used to modulate CPE levels. The method  
 can be used for diagnosing, treating and preventing type II diabetes. CPE  
 polypeptides and polynucleotides are useful for treating and diagnosing  
 type II diabetes. They are also useful for identifying compounds that can  
 modulate expression or function of CPE, and are useful for treatment and  
 diagnosis of type II diabetes. The present sequence represents human CPE,  
 which is used in the exemplification of the present invention

XX SQ Sequence 476 AA;

Query Match 30.3%; Score 1054; DB 3; Length 476;

Best Local Similarity 48.2%; Pred. No. 4.8e-94;

Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLADEALPSGLPPTFIR-----FSHSYQAQMVVRVLRRTASCAHVARTYSI 202

Db 20 GWLLGAQAQEPGAPAGRRRRRLRQEDGIGSFYHRYPELREALVSVWLQCTAISRIYTV 79

QY 203 GRSPDGRRELLVIEFSSRPGQHELMPEVVKLIIGNHNEVAGREMLIYLAQYLCSEYLLGN 262

Db 80 GRSPGGRRELLVIELSDNPGVHEPGEPEFKYIGNHNEVAGREMLIYLAQYLCSEYQKGN 139

QY 263 PRIQLRLNTTRIHLPLPSINPDGYEVAAGAGYNGWTSGRONQNLDIRNFPDLTSEY 322

Db 140 ETIVNLHSTRIHIMPSLNDPGFEKASQPGELKDWPFVGRSNAQGDIDLNRFPPDLDRIVY 199

QY 323 RLAEARGARDHI-----PIPOHYMWKGVAPETKAIMKMWOTIPFVLSASLHGGDLVVS 376

Db 200 -VNEKEGGPNHLLKNKIVDQN---TKLAPETKAVIHWMIDIPFVLSANLHGGDLVAN 255

QY 377 YPFDGSKHPOEKMFSPTPEKFKLLSRAYADVHPMMDRSENRCGN-----FLKRG 431

Db 256 YPYDETRSGAHE-YSSCPDDAIFQSLARAYSSFNPMSPDNRPPCRKNDSDSSFV--DG 312

QY 432 IINGADWYSFTGMSDFNYLHNTCFEITVELGCVKFPPEEALYTLQHNKESILNFVEIV 491

Db 313 TTNGGAWYSVPGGMQDFNYLSSNCFEITVELSCEKFPPEETLKSYPWEDNKNLSINYLEQI 372

QY 492 HRGIGKVVTDKFGKPVKNARISVKGIRHDITTPADGDYWRLLPPIGHIHIVIAQAPGYAKVI 551

Db 373 HRGVKGFVRDLQGNPIANATISVDGIDHDVTSKDGWYRLLVPGNYKLTASAPGYLAIT 432

QY 552 KKVLIIPARMKGRVDFILQ 571

Db 433 KKVAVP--YSPAAGVDFELE 450

## RESULT 9

AA23880

ID AAM23880 standard; protein; 476 AA.

XX AC AAM23880;

XX 12-OCT-2001 (first entry)

XX Human EST encoded protein SEQ ID NO: 1405.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

XX diagnostics; forensic test; gene mapping; genetic disorder; biodiversity;

XX gene therapy; nutrition.

XX Homo sapiens.

XX WO200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US002687.

XX XX

```

PR 25-JAN-2000; 2000US-00491404.
PR 17-JUL-2000; 2000US-00617746.
PR 03-AUG-2000; 2000US-00631451.
PR 15-SEP-2000; 2000US-00663870.
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
PI
XX WPI; 2001-476164/51.
DR N-PSDB; AAH98539.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use.
XX
XX Claim 20; Page 982-983; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensic, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention
XX
XX Sequence 476 AA;
SQ
Query Match 30.3%; Score 1054; DB 4; Length 476;
Best Local Similarity 48.2%; Pred. No. 4.8e-94;
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;
QY 155 GGLEADALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHVARTYSI 202
DB 20 GWLLGAEAOEPGAPAGMRRLRLQEDGISFEYHRYPELREALVSWLQCTAISRIYTV 79
QY 203 GRSGFGRRELLVIEFSSRPGQHELMPEVKLIGNHNEVAGREMLIYLAQYLCSEYLLGN 262
DB 80 GRSGFGRRELLVIELSDNPGVHEPCEPEFKYIGNHMGNEAVGRELLIFLAQYLCNEQKGN 139
QY 263 PRIQRLNTRIHLLPSINPDGVEVAAGAGYNGWTSGRQNAQNLDLNRNFPDLTSEYY 322
DB 140 ETIVNLHSTRIHIMPSLNPDGPEKAASQPCELMKDFVGRSNAQGDILNRNFPDLRIVY 199
QY 323 RLAEETRGRSDHI-----PIPOHYWGWKVAPETKAIKMWQTIPTFVLSASLHGGDLVVS 376
DB 200 -VNEKEGGPNHLLKNMKTIVDQN---TKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255
QY 377 YPFDFFSKHPQBEKMFSPTPDEKMFKLISRAYADVHPMMMDRSENRCGN-----FLKGRS 431
DB 256 YPDETRSGSAHE-YSSSPDAIFQSARAYSFNPAMSPDPNPPCKEKNDSDSFV--DG 312
QY 432 IINGADWYSTGMSDFNYLHNCFTETVBLGCVKPPPEALYTLMOHNKESLLNFVETV 491
DB 313 TTNGGAWYSPVGGWQDFNYLSSNCFEITVELSCEKPPPEETLKTWEDNKNLSIYLEQI 372
QY 492 HRGKIGVVTKFGKPVKNARISVKGIHSDITTPADGDYWLPPGHIHIVTAQAPYAKVI 551
DB 373 HRGKVGFEVRLQGNPIANATISVEGDHVDVTSAKDGYWELLIPGNYKLTASAPGYLAIT 432
QY 552 KKVIIIPARMKRAGRVDFILQ 571
DB 433 KKVAVP--YSPAGVDFELE 450
RESULT 10
ABG61920
ID ABG61920 standard; protein; 476 AA.
XX AC
XX ABG61920;
DT 15-AUG-2002 (first entry)

```

```

XX XX
DE DE
XX XX
KW KW
XX XX
OS Mammalia.
XX
XX WO200230268-A2.
XX
XX 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US032045.
XX
XX 13-OCT-2000; 2000US-00687576.
XX
XX 08-DEC-2000; 2000US-00733288.
XX
XX 08-DEC-2000; 2000US-00733742.
XX
XX 24-JAN-2001; 2001US-0263957P.
XX
XX 16-MAR-2001; 2001US-0276791P.
XX
XX 16-MAR-2001; 2001US-0276888P.
XX
XX 06-APR-2001; 2001US-0281922P.
XX
XX 24-APR-2001; 2001US-0286214P.
XX
XX 30-APR-2001; 2001US-00847046.
XX
XX 04-MAY-2001; 2001US-0288589P.
XX
XX (BOSB-) EOS BIOTECHNOLOGY INC.
PA
XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX
XX WPI; 2002-471335/50.
XX
XX N-PSDB; ABK92238.
XX
XX Detecting a prostate cancer-associated transcript in a cell in a patient,
XX useful for diagnosing prostate cancer (PC) or screening modulators of PC,
XX by determining if prostate cancer-associated genes are expressed in a
XX prostate tissue.
XX
XX Claim 27; Page 404; 436pp; English.
XX
XX The present invention relates to methods of detecting a prostate cancer-
XX associated transcript in a cell from a patient. The method comprises
XX contacting a biological sample from the patient with prostate cancer-
XX associated polynucleotides (designated PC genes) that selectively
XX hybridize to a sequence that is at least 80% identical to them. The
XX prostate cancer-associated polynucleotide sequences are differentially
XX expressed in prostate tumour tissue or in prostate cancer and are derived
XX from the tissues of various organisms such as humans or other mammals
XX (e.g. mice, sheep and dogs). The methods of the invention are useful for
XX diagnosing and treating prostate cancer in mammals. The prostate cancer-
XX associated genes are useful for diagnosing or treating prostate cancer,
XX as well as for identifying modulators of prostate cancer or agents that
XX inhibit prostate cancer. The nucleic acid sequences are particularly
XX useful in gene therapy, as a vaccine or in antisense applications.
XX ABG61800-ABG61944 represent prostate cancer-associated proteins.
XX
XX Sequence 476 AA;
SQ
Query Match 30.3%; Score 1054; DB 5; Length 476;
Best Local Similarity 48.2%; Pred. No. 4.8e-94;
Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;
QY 155 GGLEADALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHVARTYSI 202
DB 20 GWLLGAEAOEPGAPAGMRRLRLQEDGISFEYHRYPELREALVSWLQCTAISRIYTV 79
QY 203 GRSGFGRRELLVIEFSSRPGQHELMPEVKLIGNHNEVAGREMLIYLAQYLCSEYLLGN 262
DB 80 GRSGFGRRELLVIELSDNPGVHEPCEPEFKYIGNHMGNEAVGRELLIFLAQYLCNEQKGN 139
QY 263 PRIQRLNTRIHLLPSINPDGVEVAAGAGYNGWTSGRQNAQNLDLNRNFPDLTSEYY 322
DB 140 ETIVNLHSTRIHIMPSLNPDGPEKAASQPCELMKDFVGRSNAQGDILNRNFPDLRIVY 199
QY 323 RLAEETRGRSDHI-----PIPOHYWGWKVAPETKAIKMWQTIPTFVLSASLHGGDLVVS 376

```

Prostate cancer-associated protein #121.  
Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.





XX The present sequence represents carboxypeptidase E. The specification  
 CC describes a method for identifying agents which increase the propensity  
 CC of wolframin protein to associate with a carboxypeptidase E binding  
 CC partner polypeptide. The method comprises contacting both polypeptides in  
 CC the presence of a test agent, and determining whether the association of  
 CC wolframin with carboxypeptidase E is increased. Such agents can be used  
 CC to modulate wolframin protein activity

XX Sequence 476 AA;

Query Match 30.3%; Score 1054; DB 5; Length 476;  
 Best Local Similarity 48.2%; Pred. No. 4.8e-94;  
 Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHVARTYSI 202  
 DB 20 GLLGAEAEQEPGAPAGMRRLRRLQEDGIGSFYHRYPELREALVSVWLQCTAISRIYTV 79  
 QY 203 GRSPDRELLVIEFSSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGN 262  
 DB 80 GRSPDRELLVIELSNDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEYQKGN 139  
 QY 263 PRIQRLNTHRIHLLPSINPDGVEVAAGAGYNGWTSGRQNAQLDNRNFPDLTSEY 322  
 DB 140 ETIVNLTHSTRIHIMPSLNDPGFEKAASQFELKDMFVGRSNAQGDIDLRNFPDLRIYV 199  
 QY 323 RLAEATRGARSDHI-----PIPOHYWVGKVPAPETKAIMKWMQTIFFVLSASLHGGDLVWS 376  
 DB 200 -VNEKEGPNHLLKNMKKIVDQN---TKLAPETKAVIHWIMDIPFVLSANLHGGDLVAN 255  
 QY 377 YPFDKSKHPOEKMFSPTPDEKMFKLSRAYADVHPMMDRSENRCGNN-----FLKRG 431  
 DB 256 YPYDETRSGSAHE-YSSSPDDAIFQSLARAYSSFNPAFSDPNPPCKKNDSDSFV--DG 312  
 QY 432 IINGADWYFTGMSDFNYLHNCFEITVBLGCVKPPPEALYTLQHNKESLLNFVETV 491  
 DB 313 TTNGGAWYVPGGMQDNFYLSSNCFEITVBLSCFKPPPEETLKYWEDNKNLSLYLEQI 372  
 QY 492 HRGIGKGVVTDKFGKPVKNARISVKGIRHDTITAPDGDYWRLLPGHIVIAQPGYAKVI 551  
 DB 373 HRGIGKGVRLDQNPNTANATISVEGIDHVDTSAGDGYWRLLIPGNKLTASAPYLAIT 432  
 QY 552 KKVIIIPARMKRAGRVDFILQ 571  
 DB 433 KKVAVP--YSPAAGVDFELE 450

# RESULT 13

ID ADD46264  
 XX ADD46264 standard; protein; 476 AA.

AC ADD46264;

DT 29-JAN-2004. (first entry)

XX Human Protein P16870, SEQ ID NO 11939.

XX Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.

XX Homo sapiens.

XX W0203016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.  
 DR GENBANK; P16870.

XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 476 AA;

Query Match 30.3%; Score 1054; DB 7; Length 476;  
 Best Local Similarity 48.2%; Pred. No. 4.8e-94;

Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADALPSGLPPTFIR-----FSHSHYAQMVRVLRRTASCAHVARTYSI 202  
 DB 20 GLLGAEAEQEPGAPAGMRRLRRLQEDGIGSFYHRYPELREALVSVWLQCTAISRIYTV 79  
 QY 203 GRSPDRELLVIEFSSRPGQHELMPEVKLIGNIHGNEVAGREMLIYLAQYLCSEYLLGN 262  
 DB 80 GRSPDRELLVIELSNDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEYQKGN 139  
 QY 263 PRIQRLNTHRIHLLPSINPDGVEVAAGAGYNGWTSGRQNAQLDNRNFPDLTSEY 322  
 DB 140 ETIVNLTHSTRIHIMPSLNDPGFEKAASQFELKDMFVGRSNAQGDIDLRNFPDLRIYV 199  
 QY 323 RLAEATRGARSDHI-----PIPOHYWVGKVPAPETKAIMKWMQTIFFVLSASLHGGDLVWS 376  
 DB 200 -VNEKEGPNHLLKNMKKIVDQN---TKLAPETKAVIHWIMDIPFVLSANLHGGDLVAN 255  
 QY 377 YPFDKSKHPOEKMFSPTPDEKMFKLSRAYADVHPMMDRSENRCGNN-----FLKRG 431  
 DB 256 YPYDETRSGSAHE-YSSSPDDAIFQSLARAYSSFNPAFSDPNPPCKKNDSDSFV--DG 312  
 QY 432 IINGADWYFTGMSDFNYLHNCFEITVBLGCVKPPPEALYTLQHNKESLLNFVETV 491  
 DB 313 TTNGGAWYVPGGMQDNFYLSSNCFEITVBLSCFKPPPEETLKYWEDNKNLSLYLEQI 372  
 QY 492 HRGIGKGVVTDKFGKPVKNARISVKGIRHDTITAPDGDYWRLLPGHIVIAQPGYAKVI 551

Db 373 HRGKGFVRLDQGNPIANATISVEGIDHDVTSKDGWRLIPGNKLTASAPGYLAIT 432  
 QY 552-KKVIIPARMKRAGRVDFILQ 571  
 Db 433 KKVAVP--YSPAAGVDPELE 450

## RESULT 14

ADN39258  
 ID ADN39258 standard; protein; 476 AA.

AC ADN39258;  
 XX

DT 17-JUN-2004 (first entry)

DE Cancer/angiogenesis/fibrosis-related polypeptide, SEQ ID NO:576.

XX Human; differential expression; cancer; angiogenic disorder;  
 KW fibrotic disorder; psoriasis; ischaemia; heart disease; atherosclerosis;  
 KW inflammatory disease; autoimmune disease;  
 KW retinal neovascularisation syndrome; scarring; uterine fibroid;  
 KW detection; diagnosis; prognosis; drug screening; drug targeting;  
 KW wound healing; contraception; cytostatic; cardiant; immunomodulatory;  
 KW vulnery; gene therapy; vaccine.

XX Homo sapiens.

XX WO2003042661-A2.

XX 22-MAY-2003.

XX 13-NOV-2002; 2002WO-US036810.

XX 13-NOV-2001; 2001US-0350666P.

XX 21-NOV-2001; 2001US-0332464P.

XX 29-NOV-2001; 2001US-0334393P.

XX 03-DEC-2001; 2001US-0335394P.

XX 14-DEC-2001; 2001US-0340376P.

XX 08-JAN-2002; 2002US-0347211P.

XX 10-JAN-2002; 2002US-0347349P.

XX 08-FEB-2002; 2002US-035250P.

XX 13-FEB-2002; 2002US-0356714P.

XX 20-FEB-2002; 2002US-0359077P.

XX 29-MAR-2002; 2002US-0368809P.

XX 04-APR-2002; 2002US-0370110P.

XX 12-APR-2002; 2002US-0372246P.

XX 05-JUN-2002; 2002US-0386614P.

XX 16-JUL-2002; 2002US-0396839P.

XX 22-JUL-2002; 2002US-039775P.

XX 22-JUL-2002; 2002US-0397845P.

XX 09-SEP-2002; 2002US-0409450P.

XX (EOSB-) EOS BIOTECHNOLOGY INC.

XX Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;

XX Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;

XX WPI; 2003-468649/44.

XX N-PSDB; ADN39257.

XX

CC invention also relates to expression vectors and host cells comprising a  
 CC nucleic acid of the invention; antibodies which specifically bind a  
 CC polypeptide of the invention; use of such antibodies for drug targeting;  
 CC and methods of screening for modulators of activity or expression of the  
 CC polypeptides and nucleic acids. The nucleic acids, polypeptides,  
 CC antibodies and methods are useful for diagnosing, prognosing and treating  
 CC cancer and other conditions such as psoriasis, ischaemia, heart disease,  
 CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal  
 CC neovascularisation syndromes, scarring and uterine fibroids. They may  
 CC also be useful in wound healing and in contraception. The present  
 CC sequence represents a polypeptide of the invention.

XX Sequence 476 AA;

Query Match 30.3%; Score 1054; DB 7; Length 476;

Best Local Similarity 48.2%; Pred. No. 4.8e-94;

Matches 212; Conservative 67; Mismatches 129; Indels 32; Gaps 8;

QY 155 GGLEADEALPSGLPTTFIR-----PSHSYAQWVRLRTASRCAHVARTYSI 202

Db -20 GWLLGAEOEPGAPAGMRRLRRLQOEDGISFEYHRYPELREALVSVMLQCTAISRIYTV 79

QY 203 GRSFDEGRELIVIEFSRQCHHELMPEVKLIQNIHNEVAGREMLIYLAQYLCSEYLLGN 262

Db 80 GRSFEGRELLVIELSDNFGVHEPGEPEFKYIGNVHGNEAVGRELILFLAQYLCNEYQKGN 139

QY 263 PRIQLLNTYRIHLLPSINPDGYEVAAGAGYNGWTSGRQNAQLDNRNFPDLTSEY 322

Db 140 ETIVNLIHSTRIHMPSLNPDGFEKASQPGELKDFVGRSNAQGLDNRNFPDLDRIVY 199

QY 323 RLAEETRGARS DHI-----PIPOHYWGWKVPETKAIMKMQTIPFVLSASLHGGDLVVS 376

Db 200 -VNEKEGGFNHLLKMKKIVDQ--TKLAPETKAVIHIMDIPFVLSANLHGGDLVAN 255

QY 377 YPFDFSKHPQEEKMFSPTPDEKMFKLLSRAYADVHPMMMDRSENRCGN-----FLKRG 431

Db 256 YPYDETRSGSAHE-YSSSFDDAIQSLARAYSSFNPNPNDPNDKDDSSVF--DG 312

QY 432 IINGADWYSFTGCGMSDFNYLHNCPEITVELGCVKFPPEEALYTLWQNKSSLLNFVETV 491

Db 313 TTNGGAWYSVPGMQDFNYLSNCFEITVELSCEKFPPEETLKTIVWEDNKNLSYLRQI 372

QY 492 HRGIKGVVTDKFGKPVKNARISVKGIRHDITTPDGDYWRLLPGGIHIVIAQAPGYAKVI 551

Db 373 HRGVKGFVRDLQGNPIANATISVEGIDHDVTSKDGWRLIPGNKLTASAPGYLAIT 432

QY 552 KKVIIIPARMKRAGRVDFILQ 571

Db 433 KKVAVP--YSPAAGVDPELE 450

## RESULT 15

ADQ17903

ID ADQ17903 standard; protein; 476 AA.

XX AC ADQ17903;

XX DT 26-AUG-2004 (first entry)

XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 720.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX OS Homo sapiens.

XX FN WO2004048938-A2.

XX PD 10-JUN-2004.

XX PF 26-NOV-2003; 2003WO-US038193.

XX PR 26-NOV-2002; 2002US-0429739P.

XX

PT useful for diagnosing, prognosing or treating cancer, comprises detecting  
 PT a nucleic acid in a biological sample.

XX Claim 12; SEQ ID NO 576; 1385pp; English.

CC The invention relates to nucleic acids and proteins (ADN38683-ADN40064)  
 CC whose expression is upregulated or downregulated in specific cancers or  
 CC other diseases such as angiogenic or fibrotic disorders, and to methods  
 CC of determining the presence or absence of a pathological cell in a  
 CC patient by detecting a nucleic acid at least 80% identical to those of  
 CC the invention or by detecting a polypeptide of the invention. The



1000